

Sérgio Magalhães Ferreira

**ANÁLISE DE INCERTEZAS EM MODELOS GEOMÉTRICOS DE
DEPÓSITOS MINERAIS USANDO GEOESTATÍSTICA E ESQUEMA DE
PLANEJAMENTO DE EXPERIMENTOS**

Trabalho de Conclusão de Curso
(Bacharelado em Geologia)

UFRJ
Rio de Janeiro
2008



UFRJ

Sérgio Magalhães Ferreira

ANÁLISE DE INCERTEZAS EM MODELOS GEOMÉTRICOS DE DEPÓSITOS
MINERAIS USANDO GEOESTATÍSTICA E ESQUEMA DE PLANEJAMENTO DE
EXPERIMENTOS

Trabalho de Conclusão de Curso de
Graduação em Geologia do Instituto de
Geociências, Universidade Federal do Rio
de Janeiro – UFRJ, apresentado como
requisito necessário para obtenção do grau
de Bacharel em Geologia.

Orientadores:

Luis Paulo Vieira Braga

Cláudio Bettini

Rio de Janeiro
Junho de 2008

FERREIRA, Sérgio Magalhães.

Análise de Incertezas em Modelos Geométricos de Depósitos Minerais usando Geoestatística e Esquema de Planejamento de Experimentos. Sérgio Magalhães Ferreira - Rio de Janeiro: UFRJ / IGeo, 2008. xiii, 45 p. : il.; 30cm

Trabalho de Conclusão de Curso (Bacharelado em Geologia) – Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Geologia, 2008.

Orientador: Luis Paulo Vieira Braga

1. Geologia. 2. Setor de Geologia Econômica – Trabalho de Conclusão de Curso. I. Luis Paulo Vieira Braga. II. Universidade Federal do Rio de Janeiro, Instituto de Geociências, Graduação em Geologia. III. Título

Sérgio Magalhães Ferreira

ANÁLISE DE INCERTEZAS EM MODELOS GEOMÉTRICOS DE DEPÓSITOS
MINERAIS USANDO GEOESTATÍSTICA E ESQUEMA DE PLANEJAMENTO DE
EXPERIMENTOS

Trabalho de Conclusão de Curso de
Graduação em Geologia do Instituto de
Geociências, Universidade Federal do Rio
de Janeiro – UFRJ, apresentado como
requisito necessário para obtenção do grau
de Bacharel em Geologia.

Orientador:
Luis Paulo Vieira Braga
Cláudio Bettini

Aprovada em:

Por:

Orientador: Prof. Luis Paulo Vieira Braga (IM/UFRJ)

Prof. Cláudio Bettini (DEGEO/UFRJ)

Prof. Cláudio Gerheim Porto (DEGEO/UFRJ)

Prof. Santiago S. Ramirez Carvajal (IM/UFRJ)

UFRJ
Rio de Janeiro
2008

Resumo

FERREIRA, Sérgio Magalhães. Análise de Incertezas em modelos geométricos de Depósitos Minerais usando Geoestatística e Esquema de Planejamento de Experimentos. 2008. Xiii, 45 p. Trabalho de Conclusão de Curso (Bacharelado em Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

O estudo das incertezas da modelagem geométrica de um depósito mineral, nas fases iniciais de exploração, é fundamental à definição das estratégias de desenvolvimento seguintes. O uso de simulação para gerar vários modelos é bastante popular, porém, os geólogos normalmente enfrentam o problema de definir o número de simulações e decidir qual delas é a mais representativa da variabilidade geológica.

Este estudo usa o Planejamento de Experimentos e interpolador de krigagem para ajudar na definição de um modelo geométrico para um depósito mineral. Um estudo de caso sintético é apresentado. Um modelo geométrico gerado em computador é amostrado em alguns pontos. Este modelo geométrico consiste em uma superfície que pode ser considerada o topo do depósito. Então, um interpolador de krigagem é obtido da amostra, gerando um modelo aproximado do depósito sintético original. Usando o Planejamento de Experimentos um conjunto de valores para o semivariograma é gerado, bem como os modelos geométricos correspondentes. A avaliação dos modelos é feita pelo cálculo do volume do depósito.

Comparando os vários modelos gerados e os volumes correspondentes, podemos decidir, entre todos, quais mostram as características mais significantes. Este procedimento pode reduzir consideravelmente a necessidade por centenas de modelos para se conceber um modelo inicial de um depósito mineral.

Palavras-chave: Simulação Numérica com R / Depósito Mineral / Modelagem Geoestatística / Planejamento de Experimentos

Abstract

FERREIRA, Sérgio Magalhães. Uncertainty Analysis in Geometrical Modeling of Mineral Deposits using Geostatistics and Experimental Design. 2008. Xiii, 45p. Course end work (Bachelor of Geology) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

The study of the uncertainty in geometrical modelling of a mineral deposit, in the initial stages of exploration, is fundamental to the definition of the following development strategies. The use of simulation to generate several models is quite popular, however, the geologists are usually faced with the problem of defining the number of simulations and which ones are the most representatives to capture the geological variability.

This study uses experimental design and kriging interpolator to assist the definition of a geometrical model of a mineral deposit. A synthetical study case is presented. A computer generated geometrical model is sampled in some points. Then, from this sample data, a kriging interpolator is obtained, generating an approximated model of the original synthetical deposit. Using experimental design a set of values for the semivariograms are generated, as well as, the corresponding geometrical models. The evaluation of the models is done through the volume calculation of the deposit.

By comparing the several generated models and their volumes we can decide which ones show the most significant features. This procedure can reduce dramatically the need for hundreds of models to conceive an initial working model of a mineral deposit.

Keywords: Numeric Simulation with R / Ore Deposit / Geostatistics Modeling / Experimental Design

Lista de Figuras

Figura 1	Tipos de modelos de realidade possíveis de serem idealizados em pesquisa científica. Diagrama modificado de Monteiro (2005).	3
Figura 2	Amostragem em malha regular. O mapa da população de dados e o seu histograma são mostrados à direita. À esquerda estão quatro malhas de amostragem diferentes sobre a mesma população e os respectivos resultados de interpolação dos dados por krigagem ordinária. É possível perceber a criação de “artefatos” (anomalias fictícias) nas amostras com malhas mais abertas.	6
Figura 3	Etapas da análise estatística. Após a obtenção de uma amostra da população (Amostragem), calculam-se as estatísticas desta amostra que serão usadas para estimar os parâmetros da própria população (Inferência Estatística).	9
Figura 4	Relação entre as funções variograma e covariograma.	14
Figura 5	Principais propriedades da função semivariograma.	15
Figura 6	Cálculo do semivariograma experimental na direção θ . Na figura, todos os pontos na região $(h \pm \Delta h, \theta \pm \Delta \theta)$ são considerados no cálculo.	16
Figura 7	Tipos de anisotropias: Geométrica, Zonal e Mista.	17
Figura 8	Obtenção da elipse de anisotropia a partir de dois variogramas calculados em direções ortogonais entre si. Pares de variogramas são assim testados até a obtenção das direções de máxima e mínima continuidade. Normalmente a análise exploratória das anisotropias inicia-se com o estudo das anisotropias apresentadas pelos variogramas nas direções 0° e 90° e 45° e 135° .	18
Figura 9	Comportamento do semivariograma próximo à origem.	19
Figura 10	Modelos de variograma. No sentido horário a partir do canto superior esquerdo temos os gráficos dos modelos exponencial, gaussiano, Bessel, esférico, linear e pepita puro.	21
Figura 11	Modelo Matern. Dependência com o parâmetro de suavização em variogramas com patamar e alcance iguais a 1 e efeito pepita nulo. Parâmetro de suavização (κ) variando de 0.1 a 2, com passo de 0.2.	22
Figura 12	Estimativas do valor de um atributo em locais não amostrados, a partir de locais amostrados na mesma área. Dependendo do método de interpolação empregado, obtêm-se, para a mesma amostra, resultados totalmente diferentes.	24
Figura 13	Dados estatísticos da população: histograma e curva de frequência acumulada.	28
Figura 14	Definição de ângulo e razão de anisotropia usados neste trabalho.	28

Figura 15	População visada. Código de cores refere-se às cotas em relação a um nível de referência arbitrário. Os valores possuem resolução espacial (<i>pixel</i>) de 10 m e são considerados ao longo de uma área quadrada de lado igual a 1000 m.	29
Figura 16	Cálculo do volume compreendido entre a superfície do depósito e o nível de referência.	30
Figura 17	Histograma, curva de frequência acumulada e estatísticas referentes à amostra aleatória.	32
Figura 18	Amostra aleatória. Mapa de distribuição dos pontos de amostragem ao longo da região de estudo. Observam-se os <i>clusters</i> dos valores mais altos onde foram estabelecidas malhas regulares de amostragem. O tamanho e cor dos pontos representam classes de valores da variável-alvo. Os valores nos eixos Norte e Leste são expressos em metros.	32
Figura 19	Escolha da vizinhança local a partir de pontos amostrais aleatoriamente distribuídos. Em (a) o critério foi o de escolher os oito pontos de dados mais próximos do ponto a ser interpolado. Este critério resulta em um agrupamento de pontos no quadrante nordeste em detrimento dos outros. Em (b) usou-se o critério dos quadrantes, escolhendo os dois pontos mais próximos em cada quadrante. Ainda existe uma deficiência de pontos a sul. Em (c) usou-se o critério dos octantes, escolhendo o ponto mais próximo em cada octante, o que resulta em uma vizinhança de amostragem representativa em termos do gradiente dos dados. Figura modificada de Yamamoto (2001).	34
Figura 20	Semivariograma experimental com 25 pontos calculados. Visualmente podemos notar a grande continuidade na origem e a ausência de efeito pepita. Entretanto, a escolha dos valores ótimos de alcance e patamar não é tarefa tão simples quando se emprega apenas a análise visual dos pontos.	37
Figura 21	Comportamento do volume da superfície interpolada em função do valor do parâmetro de suavização (κ). A curva em cor vermelha representa o volume da superfície interpolada enquanto que a curva azul indica o volume verdadeiro. No detalhe está uma ampliação do trecho do gráfico entre os valores de κ zero e cinco. É possível ver neste gráfico que a ordem de grandeza do valor do volume e de sua taxa de variação fica fora do domínio dos dados para κ maior do que três.	38
Figura 22	Planejamento fatorial mostrando a contribuição de cada parâmetro no valor final da variável de controle. Os pontos azuis representam o valor do volume da superfície predita a partir de determinada combinação de parâmetros, enquanto que a linha vermelha indica o volume verdadeiro.	40
Figura 23	Histograma e curva de frequência acumulada referentes à superfície predita.	41
Figura 24	Superfície predita usando-se os parâmetros estabelecidos pelo esquema de planejamento de experimentos. Código de cores refere-se às cotas em relação a um nível de referência arbitrário. A gridagem usada possui resolução espacial (<i>pixel</i>) de 10 m para coincidir com a resolução da população visada.	41

- Figura 25** Comparação entre a população visada (esquerda) e a “população predita” (direita). 42
- Figura 26** Nestes mapas, o código de cores representa o módulo da diferença de cota entre a superfície original e a superfície interpolada. À esquerda, a escolha dos parâmetros do variograma foi feita do modo “tradicional”, isto é, ajuste “a sentimento” sobre o variograma com auxílio da função `fit.variogram`. À direita, a escolha foi feita pelo esquema de planejamento de experimento. Este último ajuste, de fato, minimiza as diferenças de cota. 43
- Figura 27** Tela inicial do R. 48

Lista de Tabelas

Tabela 1	Valores do semivariograma experimental obtido a partir da amostra segundo a direção N45°E e com tolerância angular de $\pm 15^\circ$. Definimos 25 “classes de distâncias” ou passos de h e distância máxima de 500 m entre dois pontos como limite para o cálculo da variância. A coluna designada por np traz a quantidade de pares de pontos que entraram no cálculo da variância para cada classe de distância. A largura do intervalo de distância (tolerância de distâncias), igual a duas vezes Δh , foi definida como sendo de 20 m.	36
Tabela 2	Valores mínimo, intermediário e máximo considerados razoáveis escolhidos para o alcance, patamar e suavização do modelo teórico de variograma.	39

Sumário

1 INTRODUÇÃO	3
1.1 MODELOS DA REALIDADE	2
1.2 POPULAÇÃO E AMOSTRA	3
1.3 TAMANHO, DOMÍNIO E SUPORTE DE UMA AMOSTRA	6
1.4 ANÁLISE ESTATÍSTICA	7
2 ANÁLISE GEOESTATÍSTICA	10
2.1 VARIÁVEIS REGIONALIZADAS	10
2.2 O VARIOGRAMA	13
2.2.1 Propriedades do Semivariograma	14
2.2.2 Modelos de Covariograma	19
2.3 KRIGAGEM	23
3 PLANEJAMENTO DE EXPERIMENTOS	26
4 METODOLOGIA	27
4.1 POPULAÇÃO, VARIÁVEL-ALVO E VARIÁVEL DE CONTROLE	27
4.2 AMOSTRAGEM	31
4.3 PROCEDIMENTO PARA A INTERPOLAÇÃO DOS DADOS	33
4.3.1 Vizinhaça local	33
4.3.2 Parâmetros a ajustar	35
5 RESULTADOS	35
5.1 SEMIVARIOGRAMA EXPERIMENTAL	35
5.2 AJUSTE POR ESQUEMA DE PLANEJAMENTO DE EXPERIMENTOS --	37
6 CONCLUSÃO	44
7 REFERÊNCIAS BIBLIOGRÁFICAS	44
APÊNDICE A: O projeto R	47
APÊNDICE B: Códigos-fonte dos programas criados neste trabalho	50

1 INTRODUÇÃO

Um dos aspectos dos trabalhos de pesquisa em geologia é o caráter de dependência espacial dos dados tratados. A variável-objetivo refere-se sempre a uma dada região da superfície da Terra. O tratamento das questões é feito por amostragens e a integração dos dados da amostra para que se conheça o todo constitui o que podemos chamar de *análise espacial de dados*.

Para se proceder à análise espacial dos dados amostrais pode-se empregar tanto a *modelagem estocástica* quanto a *interpolação*. Neste trabalho, vamos estudar a interpolação por krigagem, sobretudo no que se refere à sua sensibilidade aos seus parâmetros de ajuste. Normalmente emprega-se um procedimento iterativo por critérios de tentativa e erro para a escolha de um conjunto de parâmetros de ajuste que reproduza com certa precisão as características descritivas da amostra (média e variância, por exemplo), para posterior interpolação dos dados.

A nossa proposta neste trabalho é a de desenvolver um procedimento criterioso de escolha do melhor conjunto de parâmetros, baseado na técnica de *planejamento de experimentos*. Para isto, optamos por empregar uma superfície sintética que pretende simular um topo de depósito mineral e que, desse modo, nos permite ter total conhecimento e controle de suas propriedades. Para a simulação desta superfície usamos um algoritmo de campos gaussianos randômicos (*Gaussian Random Fields*). Todas as fases subseqüentes deste trabalho, a saber, a campanha de amostragem, a obtenção do semivariograma experimental, o cálculo do volume, o ajuste dos parâmetros do modelo teórico de semivariograma e a krigagem tiveram caráter computacional. Todos os programas são baseados na linguagem de programação **R**.

O texto que se segue está estruturado da seguinte forma: inicia-se por definir os conceitos básicos e amarrar as convenções usadas neste trabalho; em seguida, vamos apresentar os fundamentos das técnicas geoestatísticas e de planejamento de experimentos; os procedimentos empregados, os resultados e as conclusões vêm ao final. Também foram incluídos dois apêndices: um com uma apresentação da linguagem de programação **R** e outro com todos os códigos-fonte dos programas criados para a realização das tarefas que constituíram este trabalho.

1.1 MODELOS DA REALIDADE

O grau de conhecimento que podemos ter de sistemas naturais de interesse geológico nunca é total dada à virtual impossibilidade de se obter informações completas de todas as suas partes. Entretanto, freqüentemente somos capazes de obter deste sistema uma quantidade de informação que se torna suficiente, dentro de um grau de confiabilidade predefinido, para atingirmos os nossos objetivos. Neste procedimento, nós trocamos o estudo do sistema em questão pelo estudo de um MODELO representativo deste sistema. Por exemplo, pela impossibilidade de se conhecer inteiramente o comportamento mecânico dos solos, o que se faz em geologia de engenharia é substituir o problema por outro que consiste na determinação em laboratório dos “Índices de Consistência” relacionados a amostras dos solos. Estes índices são dados pelos Limites de Atterberg, obtidos por procedimentos normatizados.

Os tipos de modelos que podemos formular estão sintetizados na Figura 1 e podem ser classificados em (Landim, 2003):

- Modelo conceitual: formulação mental simplificada de um sistema natural, apresentada de forma qualitativa (diagramas, esboços) ou mesmo quantitativa que buscam definir as condições de contorno (geometria) do estudo e talvez buscar ter também alguma idéia inicial acerca da sua variabilidade interna (homogeneidade e isotropia);
- Modelo escalar: representação em escala do sistema natural (modelo cristalino de minerais, blocos-diagrama, experimentos em laboratório, mapas);
- Modelo matemático: abstração de modelos físicos em função de relações matemáticas;
- Modelo determinístico: é um modelo matemático que expressa a relação matemática entre uma variável dependente como função de uma ou mais variáveis independentes. É representado por uma expressão unívoca (a cada valor da variável dependente corresponde apenas um conjunto de valores das variáveis independentes, e vice-versa) e de solução exata;
- Modelo estatístico: é um modelo matemático que expressa a relação matemática entre uma variável dependente como função de uma ou mais variáveis

independentes, sendo pelo menos uma de caráter aleatório, isto é, cujo valor não pode ser previsto em apenas uma observação; e

- Modelo estocástico: é um modelo estatístico onde pelo menos uma variável aleatória é gerada no desenrolar do processo.

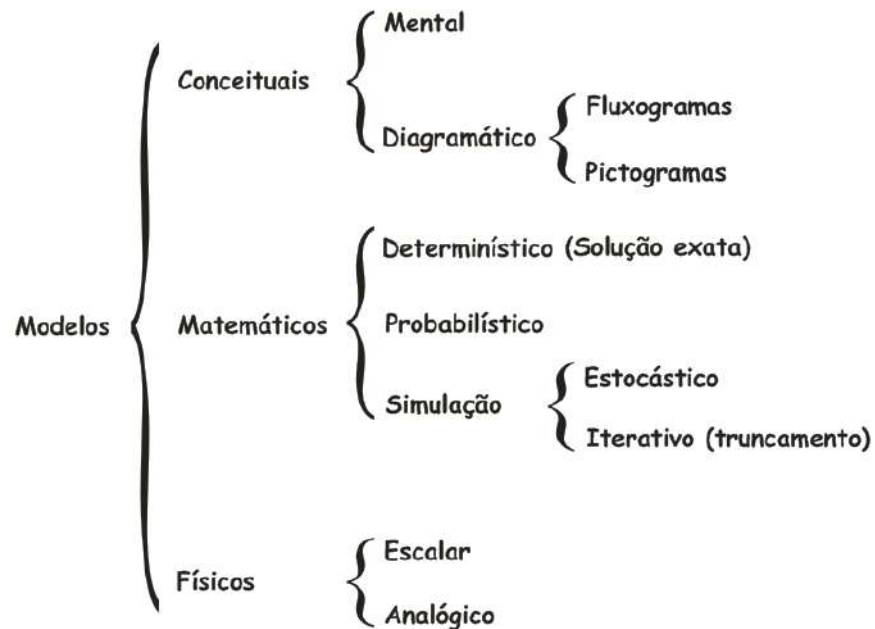


Figura 1. Tipos de modelos de realidade possíveis de serem idealizados em pesquisa científica. Diagrama modificado de Monteiro (2005).

1.2 POPULAÇÃO E AMOSTRA

Após a escolha da *região de estudo*, o passo seguinte do trabalho de investigação geológica é a determinação da *variável-alvo*, isto é, da grandeza física de onde podemos definir um subconjunto que constitui o *objetivo do estudo*. Ao conjunto de todas as medidas possíveis da variável-alvo ao longo da região de estudo denomina-se POPULAÇÃO. A *região de estudo* pode ser: uma quadrícula dentro de uma carta topográfica, uma bacia de drenagem, terrenos abaixo de uma determinada cota, um estrato específico de uma bacia sedimentar, uma fração granulométrica específica de um sistema fluvial, um corpo ígneo intrusivo, etc. Ainda podemos fazer a distinção entre *população visada*, aquela sobre a qual se está interessado e se deseja fazer inferências, e *população amostrada*, que é aquela que foi submetida a um processo de amostragem (Landim, 2003). Por exemplo, se em uma região estamos estudando certo filão mineralizado, de modo que podemos definir o teor de um mineral de minério ao longo deste filão como sendo a nossa

variável-alvo e as porções com teores acima de um determinado valor como o objetivo do estudo, então o conjunto de todas as medidas de teores possíveis ao longo do filão é a população visada.

Como raramente o conjunto de todas as observações possíveis ao longo da região de estudo acha-se disponível, os parâmetros da população visada têm de ser inferidos mediante estatísticas derivadas de amostras, as quais retornam informações de uma *população amostrada*. Portanto, AMOSTRA é o nome dado ao conjunto das observações que se tem em mãos, sendo um subconjunto da população visada. Então, o termo *observações* refere-se aos elementos da amostra enquanto que o termo *observações em potencial* refere-se aos elementos da população visada.

Perceba que o sentido da palavra *amostra* é dado em termos estatísticos e é diferente do sentido usualmente entendido em geologia ou geofísica. Em geologia, uma amostra pode ser um bloco de rocha extraído de um maciço rochoso aflorante, ou uma porção de um horizonte de solo, ou de sedimento de corrente, etc. Em geofísica, a medição de uma propriedade física de um meio geológico é uma amostra geofísica. Amostras geológicas ou geofísicas constituem indivíduos que pertencem a uma amostra em termos estatísticos. A relação entre amostra e população amostrada constitui a chamada *Inferência Estatística* (Landim, 2003), que consiste no conjunto de procedimentos e métodos usados para estimar o comportamento do conjunto de todas as observações em potencial a partir das observações disponíveis. Já a relação entre população visada e população amostrada não pode ser estabelecida por meio de métodos estatísticos, mas sim por meio do conhecimento *a priori* que se tem da população visada, dado por um modelo que agrega todo conhecimento geológico disponível acerca da região de estudo.

Ao se pensar em produzir uma amostra deve-se planejar uma campanha de amostragem o menos tendenciosa quanto possível, ou seja, que não insira nem oblitere do modelo variabilidades além daquelas que ele já possui naturalmente. Uma amostra que não seja representativa da população diz-se ENVIESADA e a sua utilização pode dar origem a interpretações erradas. Um processo de amostragem diz-se enviesado quando tende sistematicamente a selecionar elementos de alguns segmentos da população, e a não selecionar sistematicamente elementos de outros

segmentos da população. Para obtermos uma amostra de boa qualidade (que seja representativa da população visada) devemos, portanto, orientar a campanha de amostragem, sempre que possível, por um modelo bem formulado a partir das informações geológicas *a priori* da região, e garantir que as observações sejam aleatórias, isto é, que sejam coletadas imparcialmente.

O processo de amostragem dito *probabilístico* é aquele que usa mecanismos aleatórios na seleção dos elementos da amostra e onde **todos** os elementos da população têm chance de ser incluído na amostra, sendo, portanto uma amostragem sem viés. Na amostragem probabilística são utilizados com maior freqüência os seguintes tipos: amostragem casual simples, amostragem sistemática, amostragem por agrupamento (*clusters*), amostragem estratificada e amostragem hierárquica. O resultado de um processo de amostragem probabilístico é uma *amostra aleatória*. À exceção da amostragem estratificada, todos os tipos apresentados acima geram uma *amostra aleatória simples*, que é aquela em que a probabilidade de que cada observação em potencial (elemento da população visada) pertença à amostra é conhecida e igual para todas as observações em potencial (equiprováveis). No caso de amostra aleatória estratificada, a probabilidade de que cada observação em potencial pertença à amostra é conhecida, porém nem todas as observações em potencial têm igual probabilidade de pertencer à amostra.

Os elementos de uma amostra aleatória são ditos *independentes* se a amostragem foi feita com reposição, caso contrário eles são ditos *não independentes*, porém se a amostra é pequena relativamente à população então a diferença entre reposição e não reposição é atenuada. Amostras geológicas são geralmente coletadas sem reposição.

Muitas vezes em campanhas de prospecção geológica, por razões de conveniências, lança-se mão de processos de amostragem não probabilísticos, isto é, aqueles em que a probabilidade que cada observação em potencial tem de pertencer à amostra não é conhecida e o resultado não é uma amostra aleatória. No entanto, isto não implica que este processo de amostragem seja necessariamente enviesado. Um procedimento de amostragem em malha regular é um exemplo de amostragem não probabilística muito usada em campanhas de prospecção. A Figura 2 mostra o efeito que a escolha da abertura da malha de amostragem tem sobre a população inferida, dada uma mesma população visada.

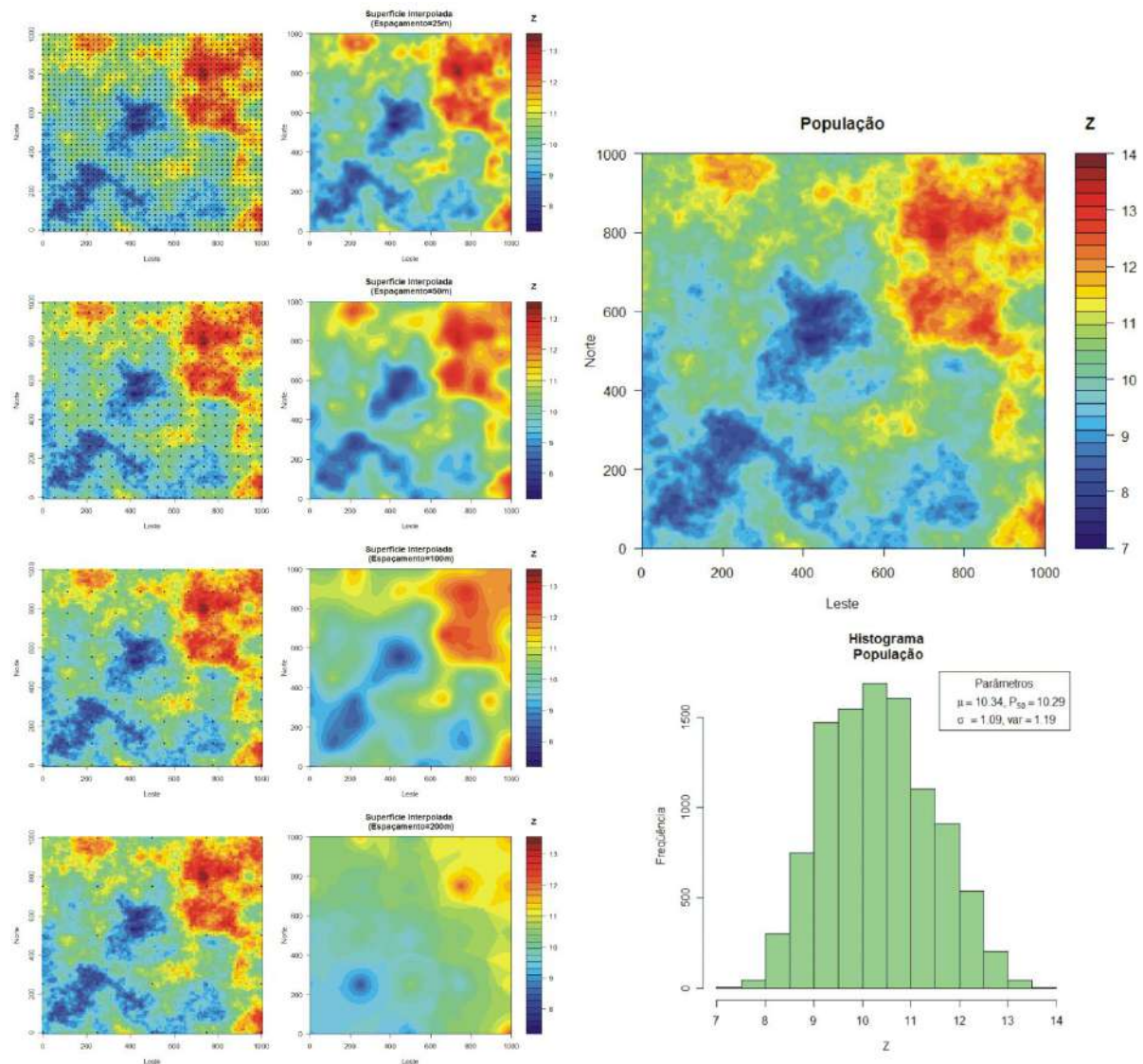


Figura 2. Amostragem em malha regular. O mapa da população de dados e o seu histograma são mostrados à direita. À esquerda estão quatro malhas de amostragem diferentes sobre a mesma população e os respectivos resultados de interpolação dos dados por krigagem ordinária. É possível perceber a criação de “artefatos” (anomalias fictícias) nas amostras com malhas mais abertas.

1.3 TAMANHO, DOMÍNIO E SUPORTE DE UMA AMOSTRA

A quantidade de observações que fazem parte de uma amostra define o TAMANHO desta amostra. Ao conjunto de todas as observações que fazem parte da amostra podemos chamar de DOMÍNIO desta amostra.

É errada a idéia de que o enviesamento de uma amostra sempre pode ser reduzido pelo aumento do tamanho da amostra. O enviesamento não tem nada a ver com o tamanho da amostra, mas sim com o processo de seleção da amostra.

Nem sempre os dados geológicos que fazem parte de uma amostra podem ser considerados pontuais ou, em outras palavras, cada observação nem sempre é tomada sobre uma região que pode ser reduzida a um ponto, mas refere-se a um comprimento ou uma área ou um volume. Assim, dados obtidos por furos de sondagem têm muito mais a ver com uma curva do que com um ponto. À unidade amostral básica sobre a qual a observação está definida chama-se SUPORTE amostral. Como veremos mais a frente, o suporte é uma característica exclusiva das variáveis regionalizadas. O suporte constitui-se na propriedade topológica das variáveis regionalizadas (Borges, 2005). Uma interessante discussão acerca dos efeitos da mudança do suporte amostral sobre a estrutura de correlação espacial dos dados pode ser apreciada em Almeida (2001) e em Almeida *et al.* (2001).

1.4 ANÁLISE ESTATÍSTICA

A *análise estatística* engloba todos os procedimentos que visam extrair informações de um conjunto de dados. Segundo Bussab & Morettin (2004), a análise estatística subdivide-se em um segmento de caráter dedutivo chamada de *Análise Exploratória de Dados* e em outro segmento de caráter indutivo chamado de *Inferência Estatística*.

A análise exploratória de dados emprega grande variedade de técnicas gráficas e quantitativas (inclusive a análise de regressão), visando maximizar a obtenção de informações ocultas na estrutura dos dados puros apresentados em tabela. A forma gráfica de se representar os dados constitui uma maneira simples e rápida de visualizar seus padrões, relações, variabilidade e tendências. Os tipos de gráficos normalmente usados para representar dados são: gráfico em barras, histogramas (gráfico de barras contíguas), gráfico de composição em setores (*pizzas*), gráfico de dispersão (pontos), gráfico ramo-e-folhas, gráfico de frequência acumulada, gráfico de quantis, gráfico de simetria e *Box plot*.

Muitas vezes estamos interessados em resumir os dados a um ou alguns valores que sejam representativos de toda a série (síntese). Para isso, empregam-se

usualmente as chamadas *medidas de tendência central*: média, mediana e moda, e as chamadas *medidas de dispersão*: variância e quantis (ou os percentis). Pode-se construir a partir daí modelos teóricos (probabilísticos) que se ajustam à distribuição dos dados. Assim, temos os modelos discretos de distribuição, tais como os modelos de distribuição Binomial, Hipergeométrico e de Poisson; e os modelos contínuos de distribuição, tais como os modelos Uniforme, Normal e Log-normal, Exponencial, Gamma e χ -quadrado, t-Student e F-Snedecor.

A Análise Exploratória de Dados possibilita a detecção de comportamentos anômalos nos dados, a descoberta de fenômenos, testar a validade das hipóteses assumidas confirmando ou não certas expectativas que se tinha sobre os dados, a escolha de modelos e o planejamento subsequente dos trabalhos.

A Inferência Estatística, como já mencionado, pretende, com base na análise de um conjunto limitado de dados (a amostra), caracterizar o todo a partir do qual estes dados foram obtidos (a população). Ela nos permite tirar conclusões e tomar decisões, indo do particular para o geral, quantificando o erro cometido nessa tomada de decisões.

A cada uma das medidas usadas para descrever características da população dá-se o nome de PARÂMETRO da população. Assim, a média e a variância das observações em potencial são parâmetros da população. A cada uma das medidas usadas para descrever características da amostra dá-se o nome de ESTATÍSTICA ou ESTIMATIVA amostral. Assim, a média, a variância e a amplitude da amostra (diferença entre o valor máximo e o valor mínimo) são estatísticas desta amostra. À estatística utilizada para estimar um determinado parâmetro chamamos ESTIMADOR do parâmetro. Quando uma amostra é recolhida, calcula-se a partir dos seus dados o valor do estimador, que dá uma estimativa do parâmetro. O conhecimento total da população se dá pelo ajuste das estatísticas a um modelo teórico probabilístico contínuo. A Figura 3 resume as etapas da análise estatística.

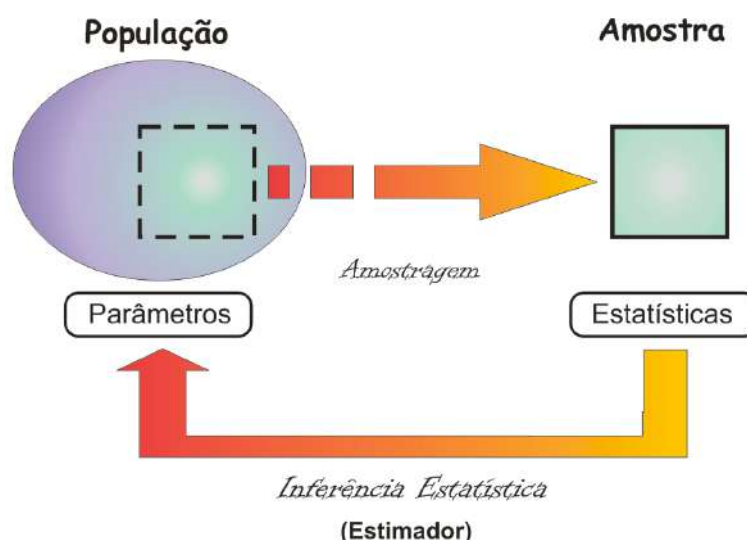


Figura 3. Etapas da análise estatística. Após a obtenção de uma amostra da população (Amostragem), calculam-se as estatísticas desta amostra que serão usadas para estimar os parâmetros da própria população (Inferência Estatística).

Se se recolher outra amostra da mesma população e com mesmo tamanho, é natural obter uma estimativa para o parâmetro diferente da primeira. Quantas amostras recolhermos, tantas serão as estimativas diferentes que podemos obter para o parâmetro. É importante chamar a atenção para o fato de que não se pode dizer qual das estimativas pontuais é melhor, já que não se conhece o valor do parâmetro a estimar. Esta variabilidade apresentada pelas estimativas é inerente à aleatoriedade da escolha da amostra e uma questão que se coloca é a de saber se o estimador que estamos considerando é um “bom” estimador ou não. Daí vem os dois problemas básicos da inferência estatística: o primeiro é chamado *teste de hipóteses* e o segundo, *estimação*.

No entanto, a tarefa de se realizar testes de hipóteses e de se buscar as propriedades dos estimadores pressupõe que os elementos da amostra são valores aleatórios independentes entre si. Assim, para uma dada amostra, a ordem dos elementos dentro do domínio é irrelevante no que se refere à estatística obtida, ou seja, as observações não são indexadas. Em outros termos, se as observações são indexadas segundo as suas posições relativas, então a inferência estatística não guarda a informação espacial das observações. Por exemplo, duas amostras com estatísticas idênticas podem possuir arranjo espacial de seus elementos totalmente diferente.

Portanto, a Análise Exploratória de Dados constitui-se em uma etapa muito importante da pesquisa geológica de recursos minerais, por exemplo, pois diz muito a respeito da população visada (ainda que de modo indireto) e também possibilita uma verificação da “qualidade” da base de dados. Já a inferência estatística falha quando se trata de dados indexados e por isso não serve para caracterizar recursos. O formalismo mais adequado para se obter uma estimativa da população, no que se refere ao arranjo espacial dos seus elementos, a partir de dados indexados no espaço é o da *geoestatística*.

2 ANÁLISE GEOESTATÍSTICA

Quando os valores dos pontos amostrais (observações) são dependentes entre si, é natural que a correlação entre eles tenda a ser maior quanto mais próximos eles estão, temporalmente e espacialmente. Podemos, com isso, indexar estes valores no espaço e/ou no tempo e formular um mapa de tendência dos dados (veremos mais a frente que esta noção está ligada ao conceito de *covariograma*). Em geoestatística a variável tempo não é importante na escala de trabalho, de modo que se costuma denotar as observações como indexadas no espaço. Nesse sentido, a amostra possui as características de uma *variável regionalizada*. Assim, GEOESTATÍSTICA é basicamente o estudo estatístico das variáveis regionalizadas.

2.1 VARIÁVEIS REGIONALIZADAS

Uma VARIÁVEL REGIONALIZADA é qualquer função numérica com uma distribuição espacial que varia de um lugar para outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função determinística (Yamamoto, 2001).

Em uma amostra definida como uma variável regionalizada, o domínio passa a possuir uma conotação geométrica. Ao comprimento do maior intervalo de amostragem dá-se o nome de CAMPO GEOMÉTRICO da amostra.

As variáveis regionalizadas possuem características que os métodos estatísticos convencionais não conseguem reconhecer. São elas:

- Seus valores são indexados pelas posições;
- São um meio termo entre variáveis determinísticas e variáveis totalmente aleatórias, onde pontos mais próximos são mais correlacionáveis do que pontos mais distantes (continuidade espacial);
- São associadas a um suporte. Isto implica que os valores em cada ponto são influenciados pela razão entre o campo geométrico e o tamanho da amostra (espaçamento médio entre os pontos de amostragem), pela forma e pela orientação em relação ao domínio; e
- Guardam informações de anisotropias e de fenômenos de transição (por exemplo, uma interface entre dois meios).

O problema que se coloca agora é como podemos fazer inferências sobre a população a partir da amostra “regionalizada”. Para isso, vamos representar a amostra como sendo a variável regionalizada $Z(\vec{X})$, em que \vec{X} representa o vetor posição de cada observação com relação a um sistema de referência, por exemplo, o sistema de coordenadas geográficas. Podemos expressar qualquer variável regionalizada como a soma de três componentes, quais sejam: uma componente estrutural determinística associada a um valor médio constante ou a uma tendência regional constante, a que chamaremos de *deriva externa (drift)* e denotaremos por $\bar{m}(\vec{X})$; uma componente estocástica, $\omega(\vec{X})$, que depende de $\bar{m}(\vec{X})$ e que possua, localmente, correlação espacial (em outras palavras, é uma variável aleatória dotada de estrutura de correlação espacial); e um termo de caráter aleatório puro, denotado por ε , com distribuição normal, média zero e variância constante, segundo Schmidt *et al.*(2002). Este termo está associado a erros de medições e outros fatores aleatórios. Assim, temos que

$$Z(\vec{X}) = \bar{m}(\vec{X}) + \omega(\vec{X}) + \varepsilon. \quad (1)$$

Então, temos que ajustar a $Z(\vec{X})$ a um modelo que descreva a sua variabilidade. A grande complexidade dos fenômenos que governam a ocorrência de $Z(\vec{X})$ torna inviável a formulação de um modelo teórico, sendo mais prático a determinação de um modelo empírico. Entretanto, como vimos no caso da inferência estatística, amostras diferentes normalmente retornam estatísticas diferentes. No caso das variáveis regionalizadas é possível fazer inferência estatística baseada em apenas uma amostra quando se considera a chamada *hipótese intrínseca*. Sejam

$Z(\vec{X})$ e $Z(\vec{X} + \vec{h})$ os valores de Z nos pontos \vec{X} e $\vec{X} + \vec{h}$, onde \vec{h} é o *vetor deslocamento* entre estes dois pontos, de modo que seu módulo, $\|\vec{h}\| = h$, é a distância entre os pontos, a hipótese intrínseca estabelece que a variável aleatória $Z(\vec{X}) - Z(\vec{X} + \vec{h})$ é *estacionária de segunda ordem*, ou seja, a média e a variância da distribuição das diferenças entre dois pontos amostrais é constante para todo o domínio e depende apenas da distância e orientação entre estes pontos (\vec{h}). Exprimindo de forma matemática fica:

$$E[Z(\vec{X}) - Z(\vec{X} + \vec{h})] = \vec{m}(\vec{h}) \quad (2)$$

e

$$VAR[Z(\vec{X}) - Z(\vec{X} + \vec{h})] = E\{[Z(\vec{X}) - Z(\vec{X} + \vec{h}) - \vec{m}(\vec{h})]^2\} \equiv 2\vec{\gamma}(\vec{h}), \quad (3)$$

onde os operadores E e VAR representam, respectivamente, a *esperança matemática* (ou *valor esperado* ou *média*) e a *variância*, que são os dois primeiros momentos centrais da variável $Z(\vec{X}) - Z(\vec{X} + \vec{h})$. O termo $\vec{m}(\vec{h})$ representa a deriva externa (*drift*) em relação ao par considerado. Nos casos mais comuns em geologia, onde não se considera a deriva, tem-se $\vec{m}(\vec{h}) = \vec{0}$. A expressão (2) representa a condição de estacionaridade de primeira ordem e diz que o valor esperado da diferença entre $Z(\vec{X})$ e $Z(\vec{X} + \vec{h})$ é a mesma para todos os pontos \vec{X} do domínio e só depende do vetor \vec{h} . A expressão (3) representa a condição de estacionaridade de segunda ordem. Para $\vec{m}(\vec{h}) = \vec{0}$, ela resulta em:

$$E\{[Z(\vec{X}) - Z(\vec{X} + \vec{h})]^2\} = 2\vec{\gamma}(\vec{h}). \quad (4)$$

No segundo membro desta expressão aparece o número 2 para descontar o fato de que o expoente no primeiro membro faz com que cada par seja considerado duas vezes. A quantidade $2\vec{\gamma}(\vec{h})$ é chamada de VARIOGRAMA e é a própria expressão da hipótese intrínseca. O variograma é definido apenas dentro do campo geométrico de $Z(\vec{X})$.

Se o domínio for definido ao longo de uma superfície (caso mais comum do emprego da análise geoestatística na geologia) ou um volume, então para cada ponto \vec{X} do domínio, a expressão (4) é definida, respectivamente, ao longo de uma

circunferência e ao longo de uma superfície esférica com centro em \vec{X} e raio h . Portanto, os modelos básicos na geoestatística supõem *estacionaridade* e também *isotropia*. Entretanto, através de transformações matemáticas convenientes, os fenômenos não estacionários e anisotrópicos também podem ser tratados.

2.2 O VARIOGRAMA

O variograma é a expressão da variabilidade de uma variável regionalizada ao longo do seu domínio. É, no entanto, preferível trabalhar com o *semivariograma*, dado por $\vec{\gamma}(\vec{h})$, ao invés do variograma, visto que este fornece duas vezes a mesma informação para cada ponto do domínio de $Z(\vec{X})$. A obtenção do semivariograma experimental a partir de uma amostra e o ajuste deste semivariograma experimental a um modelo teórico de semivariograma constitui o que se chama de ANÁLISE GEOESTATÍSTICA (por vezes referendada por *análise variográfica* ou *variografia*). O semivariograma é a ferramenta básica empregada no procedimento de interpolação de dados consagrado pela geoestatística e conhecido por *krigagem*.

Definindo:

$$\mu = E[Z(\vec{X})], \quad (5)$$

$$VAR[Z(\vec{X})] = E\{[Z(\vec{X}) - \mu]^2\}, \text{ e} \quad (6)$$

$$C(\vec{h}) = E[Z(\vec{X} + \vec{h}) \cdot Z(\vec{X})] - \mu^2, \quad (7)$$

como a *esperança* de $Z(\vec{X})$, a *variância* de $Z(\vec{X})$ e a *covariância* entre $Z(\vec{X})$ e $Z(\vec{X} + \vec{h})$, respectivamente, então, para $\vec{h} = \vec{0}$, tem-se que:

$$C(\vec{h} = \vec{0}) = VAR[Z(\vec{X})] \equiv \sigma^2, \quad (8)$$

e com isso obtém-se:

$$\vec{\gamma}(\vec{h}) = \sigma^2 - C(\vec{h}). \quad (9)$$

A expressão (9) representa a função semivariograma expressa em termos da covariância $C(\vec{h})$, dada pela expressão (7), que também é referida como FUNÇÃO

COVARIOGRAMA. A análise geoestatística busca, portanto, ajustar um modelo teórico de covariograma ao conjunto de dados $Z(\vec{X})$.

Ainda podemos acrescentar ao semivariograma uma descontinuidade na origem, usualmente chamada de EFEITO PEPITA (*nugget*) e que vamos aqui denotar por C_0 , de modo que a (9) toma a forma:

$$\hat{\gamma}(\vec{h}) = C_0 + \sigma^2 - C(\vec{h}). \quad (10)$$

As expressões (9) e (10) evidenciam o comportamento complementar entre o variograma e o covariograma, de modo que a soma entre eles é constante para todo $h \geq 0$ e igual a σ^2 (Figura 4).

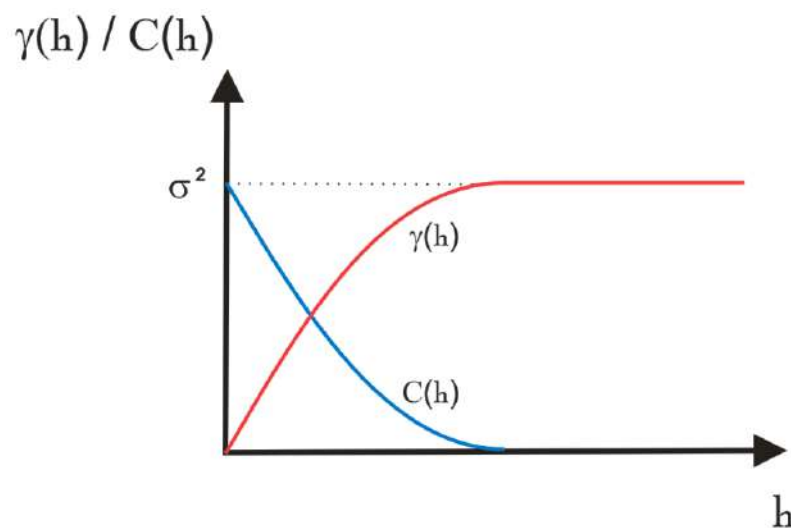


Figura 4. Relação entre as funções variograma e covariograma.

2.2.1 Propriedades do Semivariograma

Varias informações acerca de uma variável regionalizada pode ser obtida de modo simples pela análise gráfica do semivariograma. As principais propriedades do semivariograma estão representadas na Figura 5 e são as seguintes:

- Amplitude (*range*): é a distância máxima de correlação entre dois pontos. Significa que a influência mútua no valor da variável-alvo para pontos afastados por distâncias maiores do que a amplitude é nula. Define, portanto, uma *zona de influência* em torno de cada ponto do domínio. Define também uma fronteira entre o *campo estruturado* e o *campo aleatório*;

- Campo Estruturado: região no entorno de cada ponto do domínio dentro da qual os valores da variável-alvo possuem correlação espacial com o valor observado no ponto considerado. Representa o grau de continuidade espacial da variável-alvo observada localmente em cada ponto;
- Campo Aleatório: Região onde os valores da variável-alvo não possuem correlação com o valor observado em um dado ponto;
- Efeito pepita (*nugget*): é a *variância aleatória* que eventualmente possa existir na determinação do valor da variável-alvo em cada ponto. É representado por C_0 e aparece no variograma como uma descontinuidade na origem (note que o valor de $\vec{\gamma}(\vec{0}) = \vec{0}$, sempre). O efeito pepita pode ser provocado por uma micro-variabilidade intrínseca relativamente grande da variável-alvo (a exemplo da mineração de ouro, onde a inclusão de uma pepita de ouro em uma amostra é um evento aleatório); ou quando a resolução da amostragem (menor distância entre dois pontos do domínio) é maior do que a variabilidade natural da variável-alvo, resultando na ausência de correlação na escala do menor espaçamento entre pontos do domínio. Também pode ser devido a erros de medidas. Todos estes fatores **não** são mutuamente exclusivos (podem ocorrer simultaneamente);
- Patamar (*sill*): é o valor em que $\vec{\gamma}(\vec{h})$ permanece constante para $h \geq a$, onde $a = \text{amplitude}$. É a variância intrínseca de $Z(\vec{X})$, dada por σ^2 . A diferença entre a variância intrínseca e a variância aleatória dá origem a uma propriedade de $Z(\vec{X})$, a qual se dá o nome de *variância espacial (parcial sill)*.

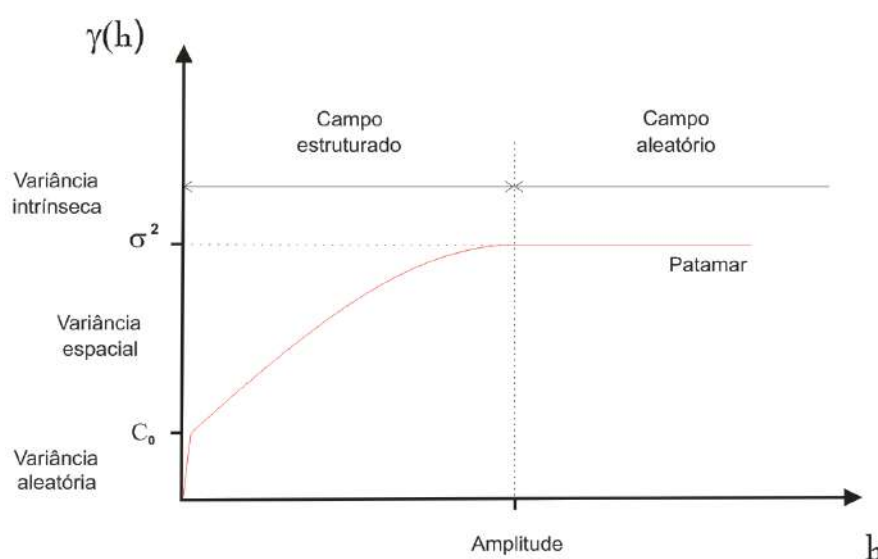


Figura 5. Principais propriedades da função semivariograma.

O estudo de semivariogramas também nos permite mapear as eventuais anisotropias no padrão dos dados. A função $\gamma(\vec{h})$ é uma função com caráter vetorial, pois, como vimos, a variabilidade expressa por ela está relacionada a uma direção específica, que no caso é a direção do vetor \vec{h} . O semivariograma direcional é, desse modo, obtido para pontos afastados de uma distância h do ponto de referência e na direção azimutal θ de \vec{h} , com tolerâncias de Δh e $\Delta\theta$, respectivamente (Figura 6). Ainda é possível se calcular um semivariograma experimental com tolerância direcional igual a 90° , isto é, que avalia todas as direções no plano horizontal. Este semivariograma é chamado de *omnidirecional* e é empregado quando se sabe de antemão que o fenômeno a ser modelado é isotrópico.

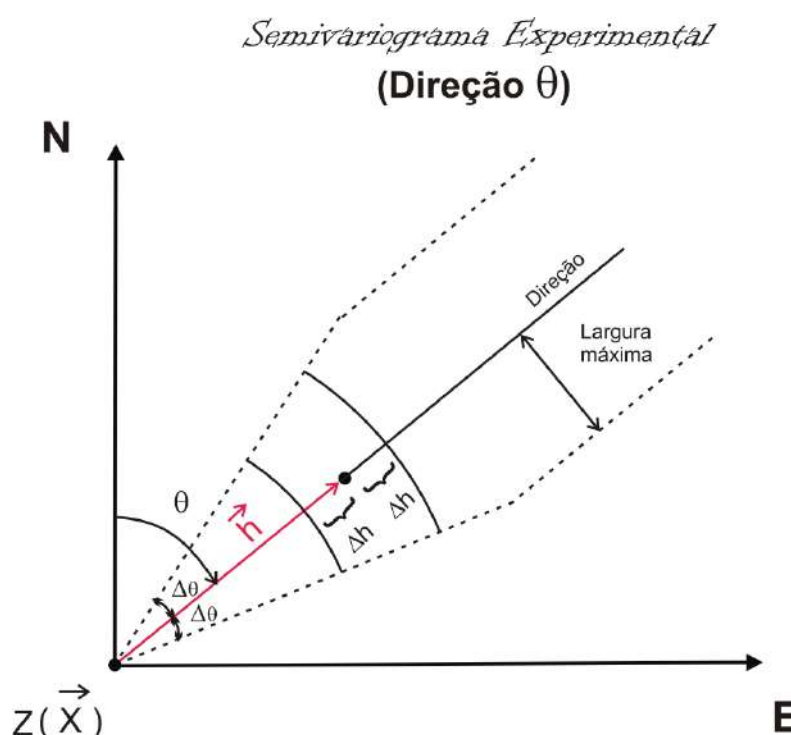


Figura 6. Cálculo do semivariograma experimental na direção θ . Na figura, todos os pontos na região $(h \pm \Delta h, \theta \pm \Delta\theta)$ são considerados no cálculo.

As anisotropias que podem ser identificadas pelo estudo do semivariograma são de três tipos: *geométrica*, quando a amplitude varia conforme as direções, mas sob um patamar constante; *zonal*, quando a amplitude permanece constante e o patamar varia de acordo com a direção; e *mista* quando tanto a amplitude quanto o patamar variam conforme as direções (Figura 7).

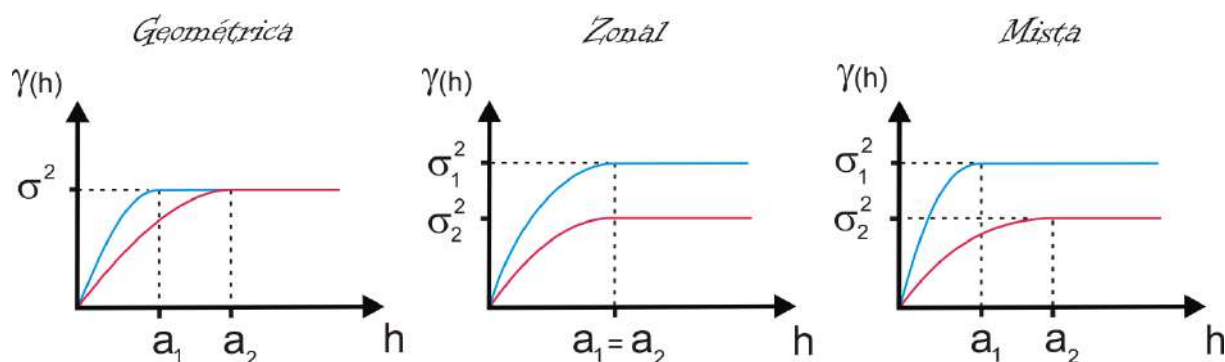


Figura 7. Tipos de anisotropias: Geométrica, Zonal e Mista.

Chamaremos de *Análise Exploratória das Anisotropias* o estudo do semivariograma experimental com a finalidade de identificar as eventuais anisotropias da variável-alvo ao longo da área de trabalho. Este estudo é feito calculando-se pares de semivariogramas em direções ortogonais entre si e comparando-se estes semivariogramas na busca de algum dos tipos de anisotropias mostrados na Figura 7. Normalmente calculam-se os semivariogramas experimentais inicialmente nas direções 0° e 90° e nas direções 45° e 135° . Caso ocorra sobreposição dos variogramas nestas quatro direções, então o fenômeno estudado é isotrópico e o variograma omnidirecional é calculado. Caso contrário, a *elipse de anisotropia* é obtida. A direção de máxima continuidade da variável-alvo é considerada a direção principal do fenômeno estudado. O semivariograma experimental calculado nesta direção é o que possui a maior amplitude ou o menor patamar e é o que serve de base para a krigagem. Caso ainda seja interessante refinar a direção principal, repete-se o procedimento anterior usando outras direções intermediárias àquelas já usadas até ser obtido, por tentativa e erro, o nível de precisão desejado. A Figura 8 mostra como se obtém a elipse de anisotropia a partir de dois semivariogramas experimentais.

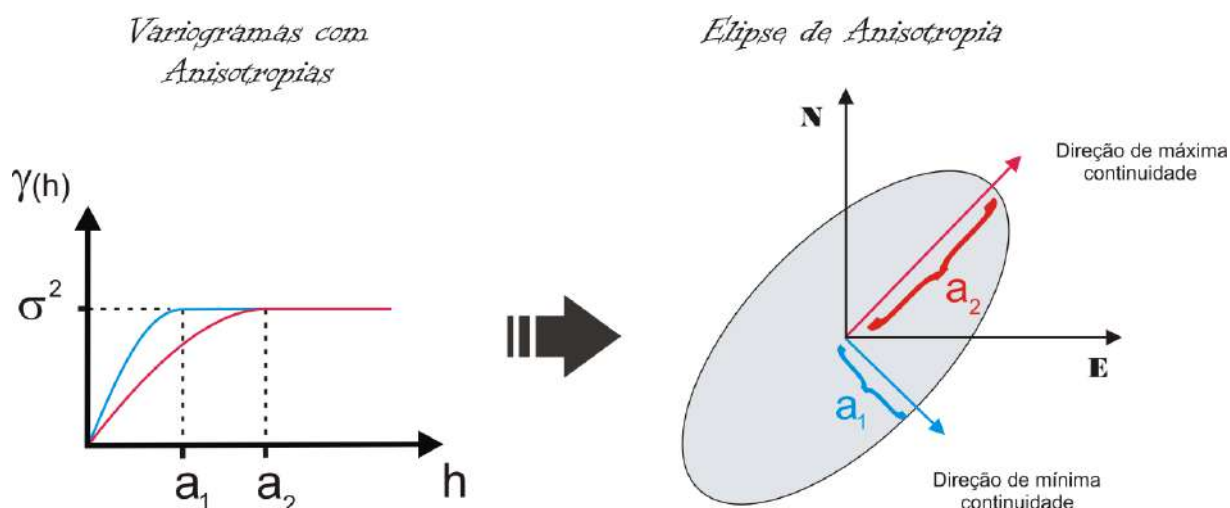


Figura 8. Obtenção da elipse de anisotropia a partir de dois variogramas calculados em direções ortogonais entre si. Pares de variogramas são assim testados até a obtenção das direções de máxima e mínima continuidade. Normalmente a análise exploratória das anisotropias inicia-se com o estudo das anisotropias apresentadas pelos variogramas nas direções 0° e 90° e 45° e 135° .

Outra propriedade de $Z(\vec{X})$ modelada pelo semivariograma é o seu *grau de continuidade*, que é dado pelo comportamento de $\gamma(\vec{h})$ próximo à origem. O ajuste deste comportamento tem a ver com a escolha de um modelo de variograma adequado. Quanto ao *tipo* de comportamento na origem, estes modelos, representados na Figura 9, podem ser:

- Parabólico: variograma descreve uma curva parabólica próximo à origem, que indica alto grau de continuidade de $Z(\vec{X})$;
- Linear: variograma exibe uma continuidade *moderada* na origem;
- Efeito pepita: variograma possui um campo estruturado, mas apresenta uma descontinuidade na origem; e
- Efeito pepita puro: variograma é não estruturado. O refinamento da malha de amostragem pode converter um variograma do tipo efeito pepita puro em um do tipo efeito pepita ou até mesmo em um do tipo linear, a não ser nos casos em que o próprio fenômeno analisado seja não estruturado (altamente aleatório).

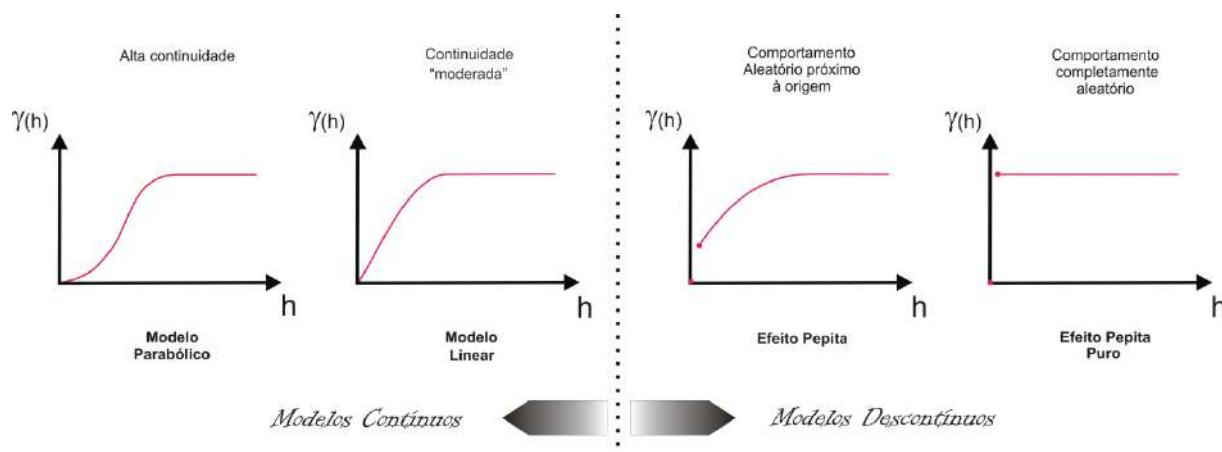


Figura 9. Comportamento do semivariograma próximo à origem.

A escolha de um modelo teórico de variograma é, portanto, ditada pela continuidade espacial (pequenas distâncias) e pelo grau de estruturação (grandes distâncias) do fenômeno analisado. Podemos substituir a tarefa de se buscar um modelo de variograma pela tarefa equivalente de se buscar um modelo de covariograma, dado que este é a expressão direta da dependência mútua entre os valores observados em dois pontos do domínio de $Z(\vec{X})$.

2.2.2 Modelos de Covariograma

Como dissemos, o variograma é peça chave da análise geoestatística. Toda informação acerca do comportamento da variável-alvo ao longo a região de estudo está concentrado no semivariograma experimental tirado da amostra. No entanto, estas são informações no estado latente, isto é, não podemos manuseá-las enquanto um ajuste a um modelo teórico não for efetuado. Esse ajuste precisa ser muito bem feito, pois todas as conclusões da pesquisa geoestatística parte dele.

Com base no padrão aparente dos pontos em um semivariograma experimental, podemos escolher dentre os modelos de covariância existente, o mais representativo. A função covariograma retorna o valor da covariância entre pares de valores da variável-alvo separados entre si de uma distância h . Esta função pode ser escrita como o produto da variância intrínseca por uma função positiva e contínua no

domínio considerado que chamaremos de *função de correlação* e denotaremos por $\rho(\vec{h})$, de modo que:

$$C(\vec{h}) = \sigma^2 \times \rho(\vec{h}). \quad (11)$$

Sendo h o módulo da distância entre os pontos e a a amplitude (distância máxima de correlação), onde $h \geq 0$ e $a > 0$, então a forma da função de correlação para alguns dos modelos mais usuais são:

- Exponencial: modelo com continuidade moderada na origem e que atinge o patamar assintoticamente.

$$\rho(\vec{h}) = e^{-(h/a)} \quad (12)$$

- Gauss: modelo com patamar bem definido e com alta continuidade na origem.

$$\rho(\vec{h}) = e^{-(h/a)^2} \quad (13)$$

- Bessel: modelo com patamar bem definido e com continuidade moderada a alta na origem. Pode ser de dois subtipos: kBessel e jBessel, sendo $\rho(\vec{h})$ dado, respectivamente, por:

$$\rho(\vec{h}) = \frac{2^\alpha \Gamma(\alpha+1)}{(h/a)^\alpha} J_\alpha \left(\frac{h}{a} \right), \text{ e} \quad (14a)$$

$$\rho(\vec{h}) = \frac{(h/a)^\alpha}{2^{\alpha-1} \Gamma(\alpha)} K_{-\alpha} \left(\frac{h}{a} \right), \quad (14b)$$

onde $\alpha > 0$ na (14a) e $\alpha > \frac{d}{2} - 1$ (d é a dimensão topológica do domínio) na (14b). $K_{-\alpha}$ é a função modificada de Bessel do terceiro tipo de ordem $-\alpha$ e J_α é a função de Bessel de ordem α .

- Esférico: modelo com patamar bem definido e com continuidade moderada na origem.

$$\rho(\vec{h}) = 1 - \frac{3}{2} \left(\frac{h}{a} \right) + \frac{1}{2} \left(\frac{h}{a} \right)^2, \text{ para } h < a, \text{ e} \quad (15a)$$

$$\rho(\vec{h}) = 0, \quad \text{para } h \geq a \quad (15b)$$

- Linear: modelo com continuidade moderada na origem e que não atinge um patamar. Portanto, não se trata de um fenômeno estacionário de segunda ordem e a função de correlação (e o covariograma) não são definidos. No entanto, o variograma pode ser definido como:

$$\vec{\gamma}(\vec{h}) = C_0 + \beta h, \text{ para } \beta > 0 \text{ e não muito maior que } 2. \quad (16)$$

- Pepita puro: modelo que se ajusta a fenômenos não estruturados.

$$\rho(\vec{h}) = k, \text{ onde } k = \text{constante.} \quad (17)$$

A Figura 10 apresenta a forma gráfica destes modelos. Vale ressaltar que nos modelos com continuidade na origem podemos “retirar” esta continuidade acrescentando um termo de efeito pepita, de acordo com a (10). Também é possível mesclar mais de um modelo, caso isto forneça um melhor ajuste e esteja em conformidade com o modelo prévio que se tem do fenômeno em estudo. O modelo *duplo esférico* aplicado a uma região com anisotropia zonal ou mista é um exemplo do emprego de modelos mesclados de covariograma.

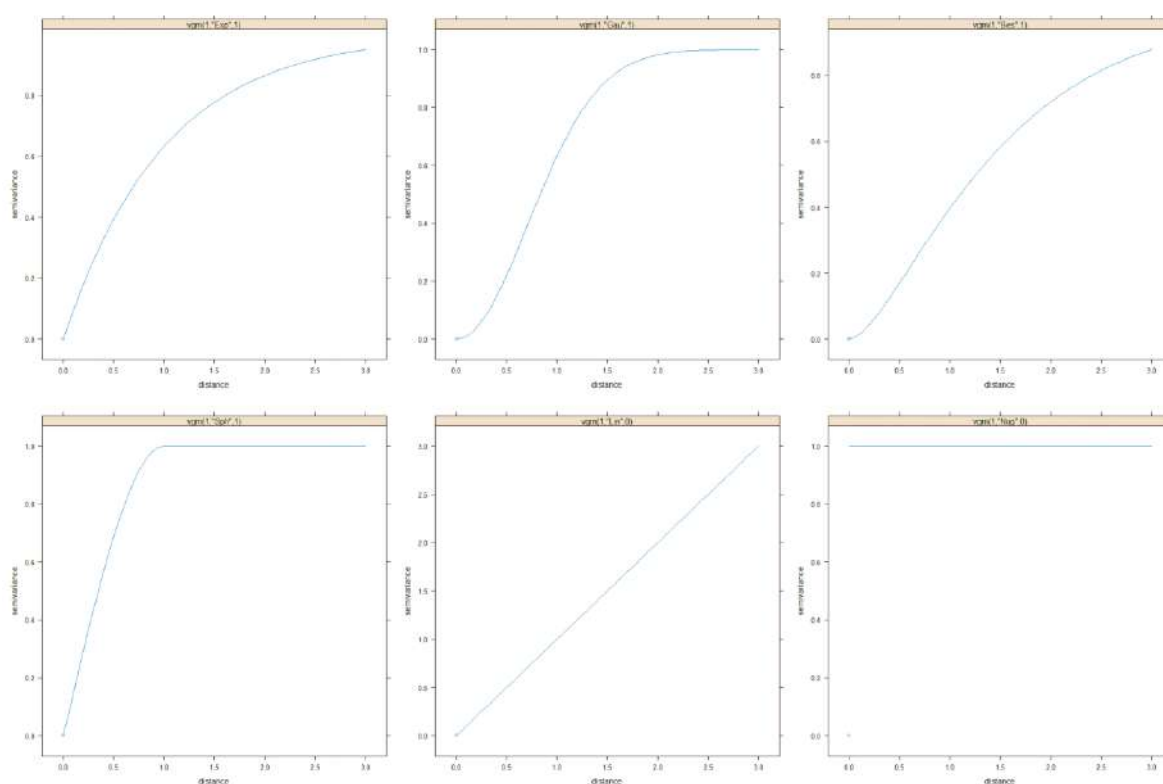


Figura 10. Modelos de variograma. Da esquerda para a direita e de cima para baixo temos os gráficos dos modelos exponencial, gaussiano, Bessel, esférico, linear e pepita puro.

Outros modelos comuns de variograma são: *Estável*, *Gamma*, *Potência*, *Seno Cardinal*, *Cauchy*, etc. Neste trabalho, vamos usar o modelo de variograma conhecido por *Matern*. Neste modelo, para um dado valor κ , que é chamado de *parâmetro de suavização* (*smoothness parameter*), sendo que $\kappa > 0$, a função de correlação é dada por:

$$\rho(\vec{h}) = \left(\frac{(h/a)^\kappa}{2^{\kappa-1}\Gamma(\kappa)} \right) K_\kappa \left(\frac{h}{a} \right), \quad (18)$$

onde K_κ é a função modificada de Bessel do terceiro tipo de ordem κ . O modelo *Matern* equivale, portanto, ao modelo *kBessel*. Este modelo foi escolhido para o presente trabalho, pois a combinação dos valores de seus parâmetros, a saber: parâmetro de suavização (κ) e parâmetro de forma (a), permite a obtenção de inúmeros “formatos” distintos para a curva. A Figura 11 mostra como a estrutura de um variograma modelado por Matern se altera como função do parâmetro de suavização.

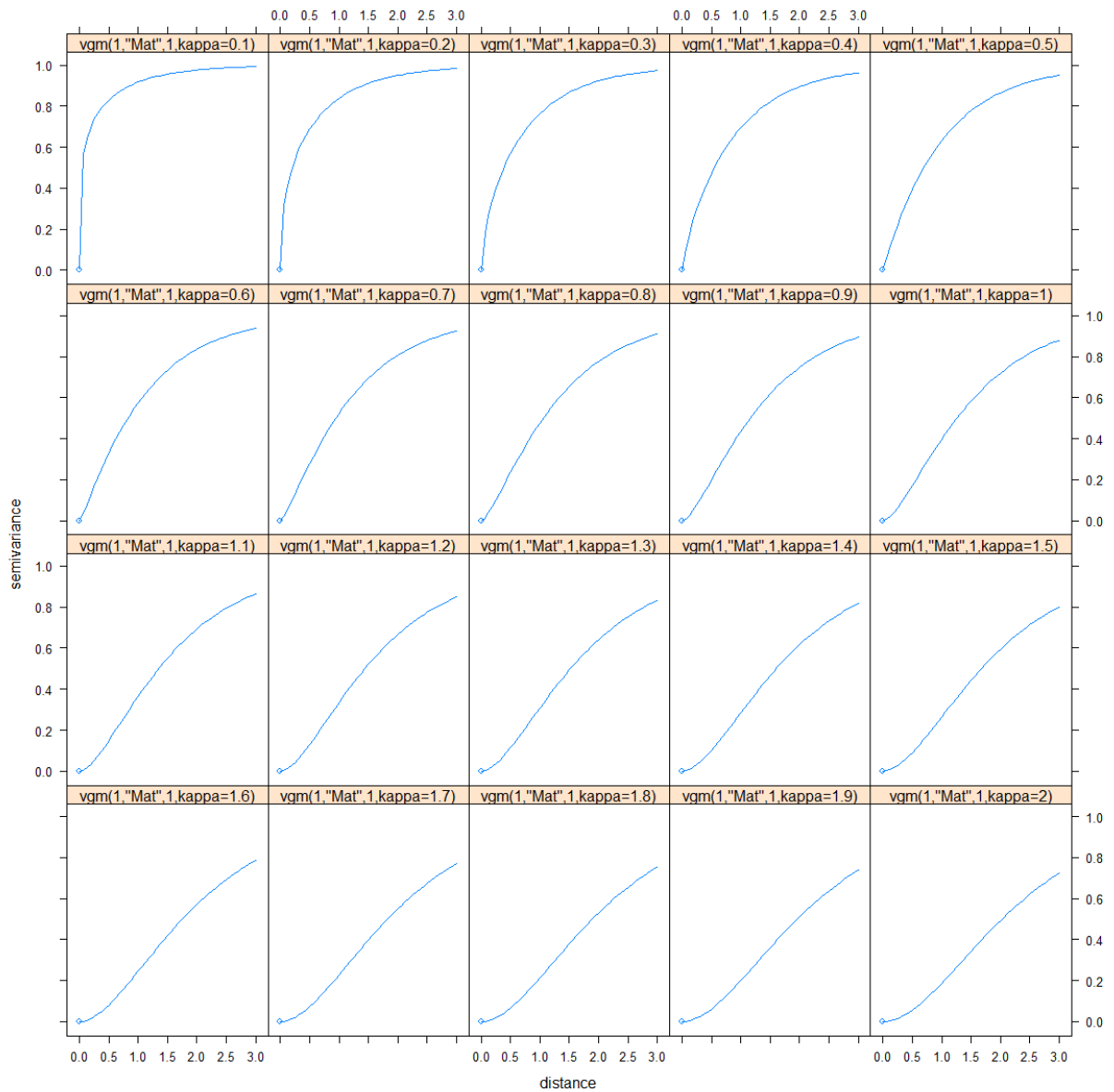


Figura 11. Modelo Matern. Dependência com o parâmetro de suavização em variogramas com patamar e alcance iguais a 1 e efeito pepita nulo. Parâmetro de suavização (κ) variando de 0.1 a 2, com passo de 0.2.

Podemos notar que todos os modelos de covariograma são modelos *paramétricos* de dependência espacial, o que significa dizer que cada modelo é representado por uma família de expressões em função dos parâmetros estabelecidos. Então, para o ajuste do semivariograma, devemos, após a escolha do modelo, proceder à tarefa de ajustar os valores dos parâmetros deste modelo.

2.3 KRIGAGEM

É uma técnica geoestatística de interpolação de dados regionalizados que fornece os melhores estimadores lineares não enviesados (*Best Linear Unbiased Estimators* – B.L.U.E.). O termo *krigagem* é derivado do nome de Daniel G. Krige, engenheiro de minas sul africano que foi o pioneiro a introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas de ouro em Witwatersrand, na África do Sul. Estas médias móveis são ponderadas por valores obtidos da análise variográfica. Esta é a principal diferença entre a krigagem e os outros métodos de interpolação: a estrutura espacial da variabilidade da variável-alvo é previamente determinada via análise variográfica. É uma interpolação de caráter local, no sentido de que para cada ponto não amostrado a ser interpolado é considerado apenas a região no entorno deste ponto. Esta região corresponde ao campo estruturado do semivariograma.

Outras características interessantes da interpolação por krigagem em relação aos outros métodos de interpolação são: a krigagem é capaz de modelar bem tanto tendências regionais como anomalias locais; a krigagem é capaz de “mapear” os erros de estimativa ao longo da área estudada através da *variância de interpolação*; e com a krigagem é possível determinar a melhor configuração espacial dos pontos amostrais através da *variância de krigagem*. A Figura 12 ilustra as diferentes superfícies obtidas por seis métodos de estimativas de caráter local referentes à mesma amostra.

Krigagem é o nome genérico dado ao processo de estimativa baseado na Teoria das Variáveis Regionalizadas. De fato, existem vários tipos de krigagem. Alguns exemplos de tipos de krigagem existentes são: krigagem simples, krigagem ordinária, krigagem pontual, krigagem de bloco, krigagem indicativa, krigagem disjuntiva, krigagem da média, krigagem transitiva, krigagem universal e co-

krigagem. As formas mais usuais são a krigagem simples e a krigagem ordinária. A krigagem simples é utilizada quando a média é assumida como estatisticamente constante para toda a área e a krigagem ordinária, por sua vez, considera a média flutuante ou móvel por toda a área. Neste trabalho empregamos a krigagem ordinária. As equações (23) representam a matriz da krigagem ordinária.

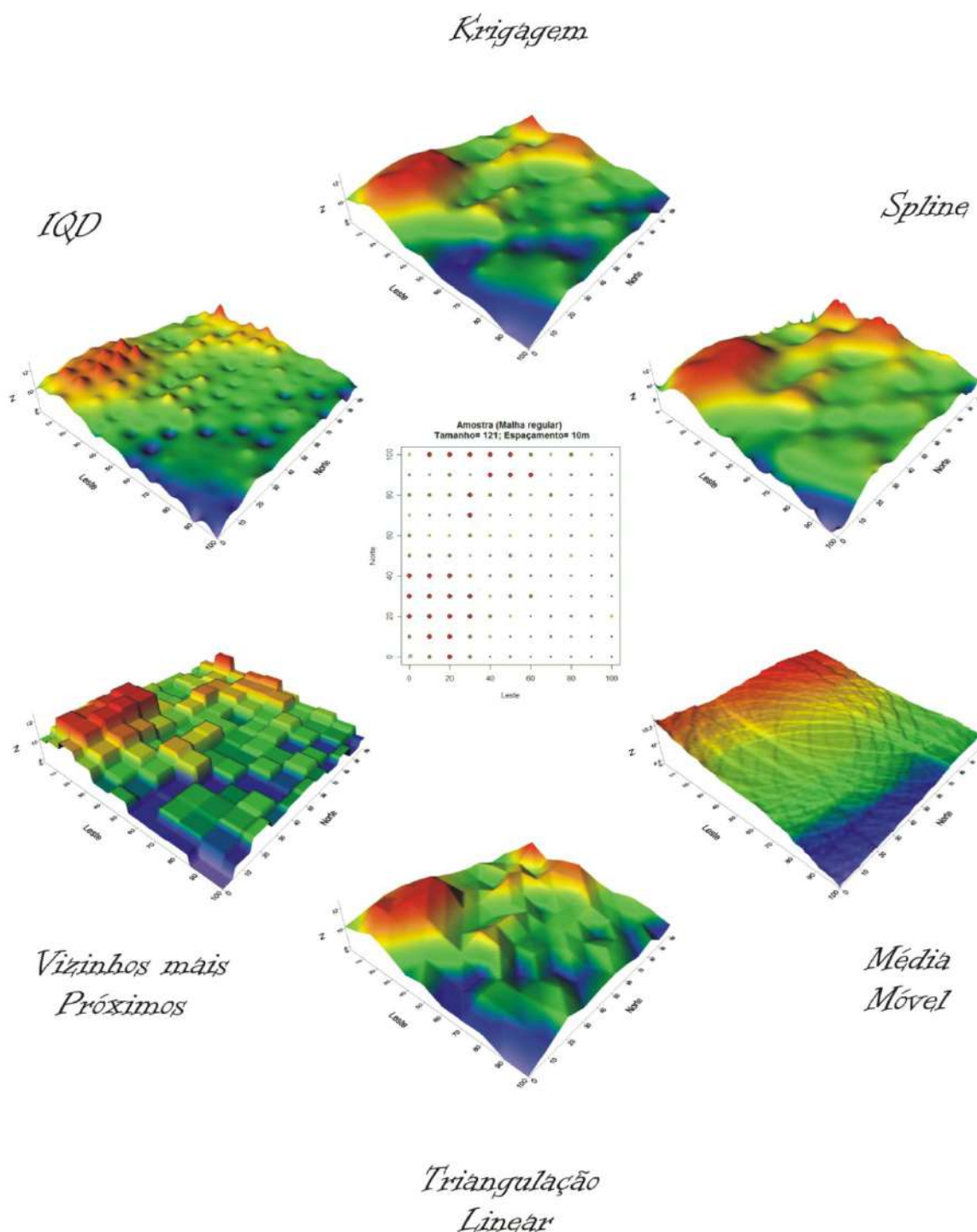


Figura 12. Estimativas do valor de um atributo em locais não amostrados, a partir de locais amostrados na mesma área. Dependendo do método de interpolação empregado, obtêm-se, para a mesma amostra, resultados totalmente diferentes.

Seja $Z^*(\vec{X}_0)$ uma estimativa do valor desconhecido de Z no ponto \vec{X}_0 , não amostrado (geralmente escolhe-se os pontos \vec{X}_0 de modo a definir uma malha regular). Podemos obter o estimador $Z^*(\vec{X}_0)$ como uma combinação linear dos dados disponíveis, de modo que:

$$Z^*(\vec{X}_0) = \sum_{i=1}^n \lambda_i \times Z(\vec{X}_i) , \quad (19)$$

onde os λ_i são os ponderadores de krigagem. Para que o estimador $Z^*(\vec{X}_0)$ seja não enviesado, basta impor que as médias real e estimada sejam iguais, ou seja, basta garantir que:

$$E[Z(\vec{X}_0) - Z^*(\vec{X}_0)] = 0 . \quad (20)$$

Com isso, a condição de não enviesamento para $Z^*(\vec{X}_0)$ fica:

$$\sum_{i=1}^n \lambda_i = 1 . \quad (21)$$

Sendo a variância do erro da krigagem dada pela equação:

$$\sigma_E^2 = VAR[Z(\vec{X}_0) - Z^*(\vec{X}_0)] , \quad (22)$$

então a minimização da variância do erro sujeita à condição de não viés resulta nas equações de krigagem (nas variáveis λ_i), dadas por:

$$\sum_j \sum_{i=1}^n \lambda_j C(\vec{X}_i - \vec{X}_j) - \xi = C(\vec{X}_0 - \vec{X}_i) , \text{ e} \quad (23a)$$

$$\sum_j \lambda_j = 1, \quad (23b)$$

onde ξ é o multiplicador de Lagrange do sistema, \vec{X}_i e \vec{X}_j são coordenadas de dois pontos amostrais e $C(\cdot)$ é a covariância espacial, que pode ser substituída pelo variograma correspondente. As expressões (23a) e (23b), expressas em termos matriciais, ficam:

$$\begin{bmatrix} C(x_1 - x_1) & C(x_1 - x_2) & \dots & C(x_1 - x_n) & 1 \\ C(x_2 - x_1) & C(x_2 - x_2) & \dots & C(x_2 - x_n) & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ C(x_n - x_1) & C(x_n - x_2) & \dots & C(x_n - x_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \xi \end{bmatrix} = \begin{bmatrix} C(x_0 - x_1) \\ C(x_0 - x_2) \\ \vdots \\ C(x_0 - x_n) \\ 1 \end{bmatrix} \quad (23)$$

A krigagem é, portanto, um interpolador linear (assume um gradiente contínuo e linear entre os pontos), aproximado (valores estimados não coincidem necessariamente com valores amostrados, mas as tendências globais dos dados são respeitadas), local, estocástico (incorpora o conceito de aleatoriedade, permitindo o cálculo de incerteza dos valores estimados), gradual (gera uma superfície contínua e suave), não enviesado e com variância mínima do erro.

Para maior detalhamento do formalismo matemático da teoria das variáveis regionalizadas, sem perder de vista a intuição geológica subjacente, recomendo a leitura de Cressie (1993), David (1982) e Davis (1986).

3 PLANEJAMENTO DE EXPERIMENTOS

Os esquemas de planejamento de experimentos são empregáveis em qualquer área das ciências onde se está interessado em saber quais variáveis são importantes em algum estudo que se esteja realizando, assim como os limites inferior e superior pertinentes destas variáveis (John, 1977). É necessário escolher, de acordo com a conveniência do pesquisador, limites inferior e superior *a priori* de valores. Corridas experimentais devem ser realizadas a fim de se obter valores para uma variável de referência para cada combinação de valores das variáveis experimentadas. A partir destes resultados, técnicas estatísticas devem ser utilizadas de modo a se concluir algo em relação à dependência da variável de referência com as variáveis analisadas. O planejamento de experimentos é uma técnica, que atualmente vem sendo usada em grande escala. Através dele, pesquisadores podem determinar as variáveis que exercem maior influência no desempenho de um determinado processo, tendo como resultado:

- Definição das variáveis importantes em um processo;
- Redução da variação do processo e melhor concordância entre os valores nominais obtidos e os valores pretendidos;
- Redução do tempo do processo;
- Redução de custo operacional; e
- Melhoria no rendimento do processo.

As variáveis que vamos analisar neste trabalho são os parâmetros de ajuste do semivariograma, quais sejam, amplitude, patamar e κ (parâmetro de suavização); e a variável de referência será o volume do depósito.

4 METODOLOGIA

4.1 POPULAÇÃO, VARIÁVEL-ALVO E VARIÁVEL DE CONTROLE

O primeiro passo do nosso trabalho foi gerar a população de dados sobre a qual vamos basear todas as tarefas subseqüentes. Queremos que esta população represente a geometria de um topo estratigráfico de um depósito mineral. A nossa variável-alvo considerada é a cota dos pontos da superfície do depósito em relação a um nível de referência. A simulação foi feita com base em um algoritmo de campos gaussianos randômicos (*Gaussian Random Fields*), usando a função `grf` do pacote *geoR* (Ribeiro & Diggle, 2001) do ambiente de programação **R**. O código-fonte do programa que gerou esta população é apresentado no Apêndice B. Uma introdução à teoria dos campos gaussianos randômicos pode ser encontrada em Harvard (2005).

As características definidas para esta população foram as seguintes:

- Malha regular de 101x101 pontos com total de 10201 pontos. Consideramos a escala que leva o lado da malha a ter 1000m de comprimento e 10m de espaçamento entre os pontos. Orientação segundo os eixos Norte e Leste;
- Os valores (população visada) representam cotas em metros, com média em 9,22 m, mediana em 9,24 m e variância igual a 0,73 m² (Figura 13);
- Os parâmetros “sugeridos” do variograma são: efeito pepita zero, patamar igual a 1 e alcance igual a 250 m; e
- Anisotropia regional. A direção de maior continuidade é N45°E e a razão de anisotropia é igual a 2 (veja na Figura 14 a definição destes termos).

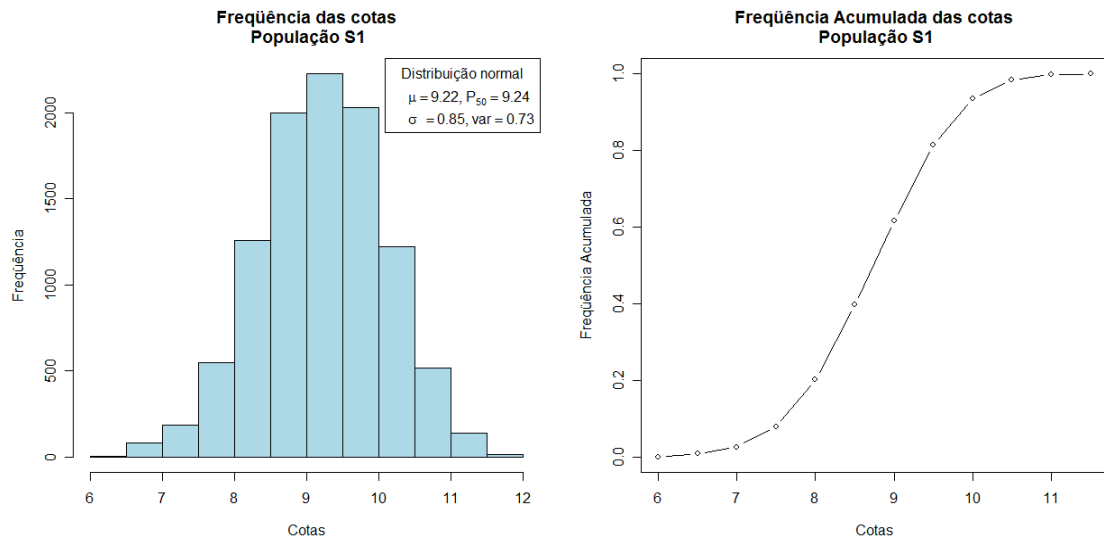


Figura 13. Dados estatísticos da população: histograma e curva de frequência acumulada.

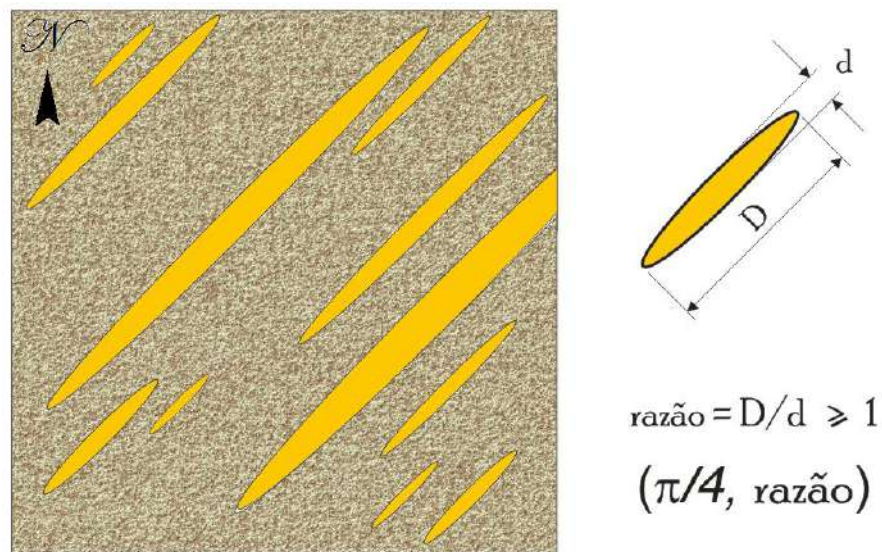


Figura 14. Definição de ângulo e razão de anisotropia usados neste trabalho.

A intenção não foi a de recriar um depósito específico, mas sim características gerais de uma superfície natural. Neste trabalho escolhemos trabalhar com **cotas** como *variável-alvo* e o **volume** como *variável de controle*, no entanto a nossa intenção é a de que o procedimento que vamos estabelecer valha para qualquer tipo de depósito e para variáveis-alvos de qualquer natureza. Para a compreensão das propriedades características de depósitos minerais, recorreremos à leitura de Annels

(1991), Arantes & Mackenzie (1995), Chaves *et al.* (1988), Guilbert & Park (1997), Guimarães (2002), Monteiro (2005) e Schobbenhaus & Silva Coelho (2001). A Figura 15 representa a nossa população de dados.

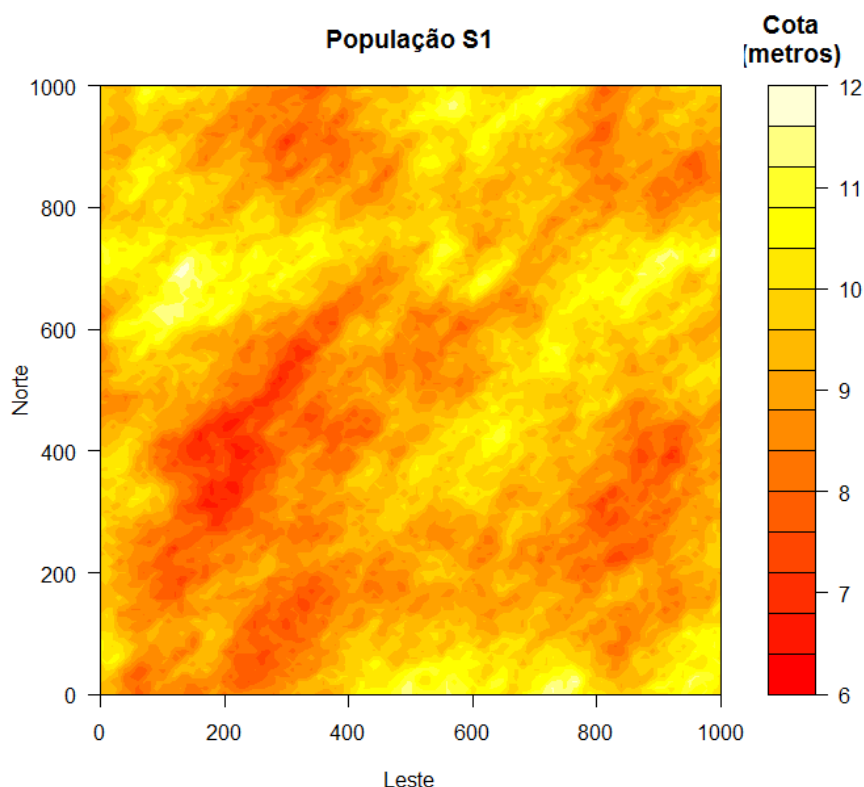


Figura 15. População visada. Código de cores refere-se às cotas em relação a um nível de referência arbitrário. Os valores possuem resolução espacial (*pixel*) de 10 m e são considerados ao longo de uma área quadrada de lado igual a 1000 m.

Para o cálculo do volume criamos uma função que pode ser aplicada a qualquer objeto expresso na forma de matriz, onde os elementos representem os valores das cotas nos nós de uma malha regular. Se no lugar de cotas tivermos teores de minérios, então o resultado desta função será a tonelagem total do minério na área considerada.

Para efetuar os cálculos, usamos a propriedade do volume de um tronco de prisma quadrangular (base quadrada) reto (arestas perpendiculares à base), pelo que o volume é dado pelo produto entre a média aritmética dos comprimentos das arestas e a área da base. A Figura 16 representa o cálculo do volume.

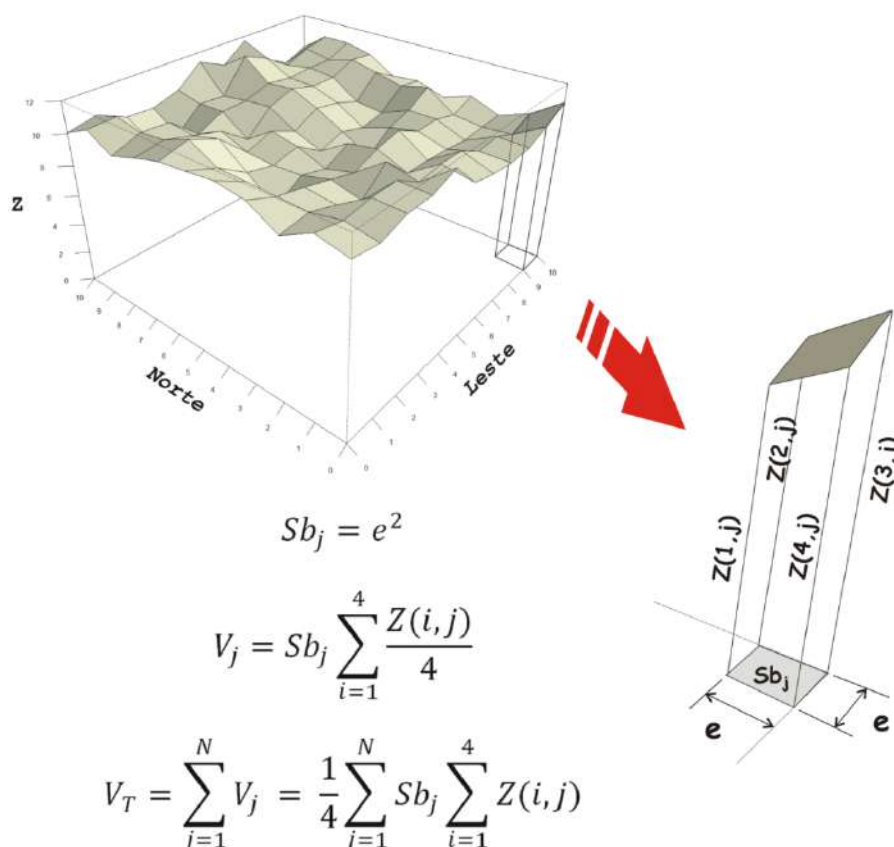


Figura 16. Cálculo do volume compreendido entre a superfície do depósito e o nível de referência.

As linhas abaixo representam o código-fonte da função empregada para o cálculo do volume, onde os argumentos **z** e **e** são, respectivamente, a matriz com os dados da variável-alvo e o comprimento do lado da base de cada prisma, que é a resolução espacial (*pixel*) da população, e vale 10m.

```
vol<- function(z, e=10){
  nrow<- nrow(z)
  ncol<- ncol(z)
  Sb<- e^2

  #Cálculo da matriz m2
  v1<- c(1,rep(2,nrow-2),1)
  v2<- 2*v1
  v<- c(v1,rep(v2,ncol-2),v1)
  m2<- matrix(v,nrow,ncol)

  #Cálculo do volume
  m<- z*m2
  a<- sum(m)
  volume <- (Sb * a)/4
  volume
}
```

4.2 AMOSTRAGEM

A partir da população sintética o passo seguinte foi simular uma campanha de amostragem ao longo do terreno e gerar, a partir daí, a base de dados ou a amostra para as etapas subseqüentes do trabalho. A intenção é a de reproduzir o mais fielmente quanto possível as técnicas de amostragem usadas em uma campanha real de coleta de dados sobre uma área onde o conhecimento geológico *a priori* é mínimo. Assim, os locais de coleta de dados foram determinados seguindo uma distribuição aleatória e nas regiões onde se detectou valores anômalos da variável-alvo foi estabelecido uma malha regular fina de pontos de coleta de dados. Estes *clusters* de refinamento da amostragem, como já discutimos, não insere à base de dados tendências além das já existentes na população visada. Trata-se de um procedimento usado para se avaliar o potencial econômico das regiões onde os valores anômalos são tais que despertam o interesse justificam um maior investimento em pesquisa. Estes *clusters* podem representar também as malhas empregadas no estudo de avaliação de reservas (cubagem). A Figura 17 mostra o histograma e a curva de frequência acumulada da amostra. A Figura 18 representa o mapa de distribuição dos pontos de coleta de dados. No Apêndice B é apresentado o código do programa que gerou esta amostra. Também apresento o código de um programa que extrai da população uma amostra definida por malha regular de amostragem com abertura definida pelo programador.

Os parâmetros caracterizadores desta amostra são os seguintes:

- Tamanho igual a 302 pontos;
- Média dos valores igual a 9,88 m, mediana em 9,75 m e variância de 1,50 m²;
- Malhas regulares com abertura de 10 m nas regiões com altos valores.

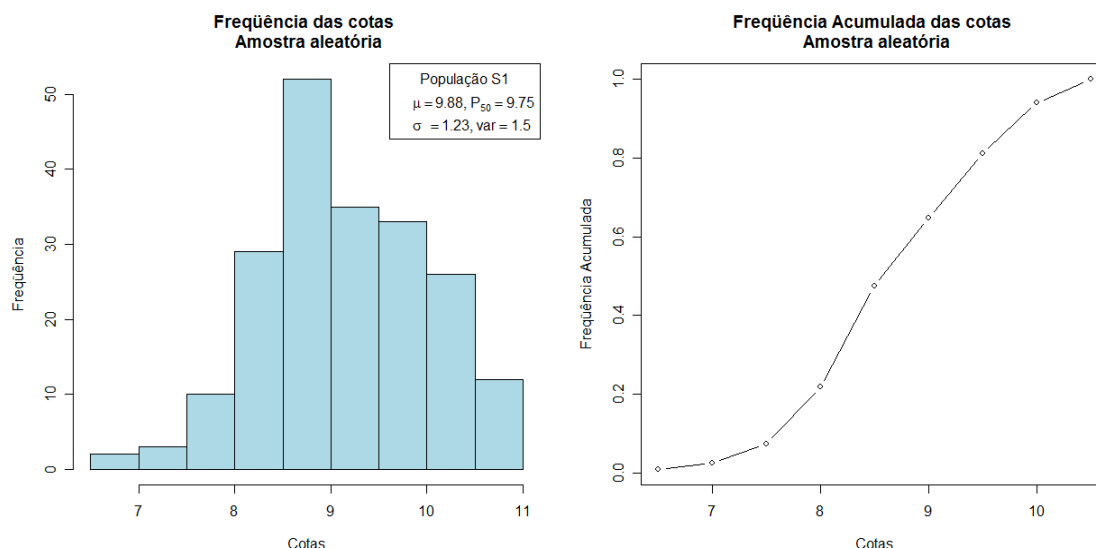


Figura 17. Histograma, curva de frequência acumulada e estatísticas referentes à amostra aleatória.

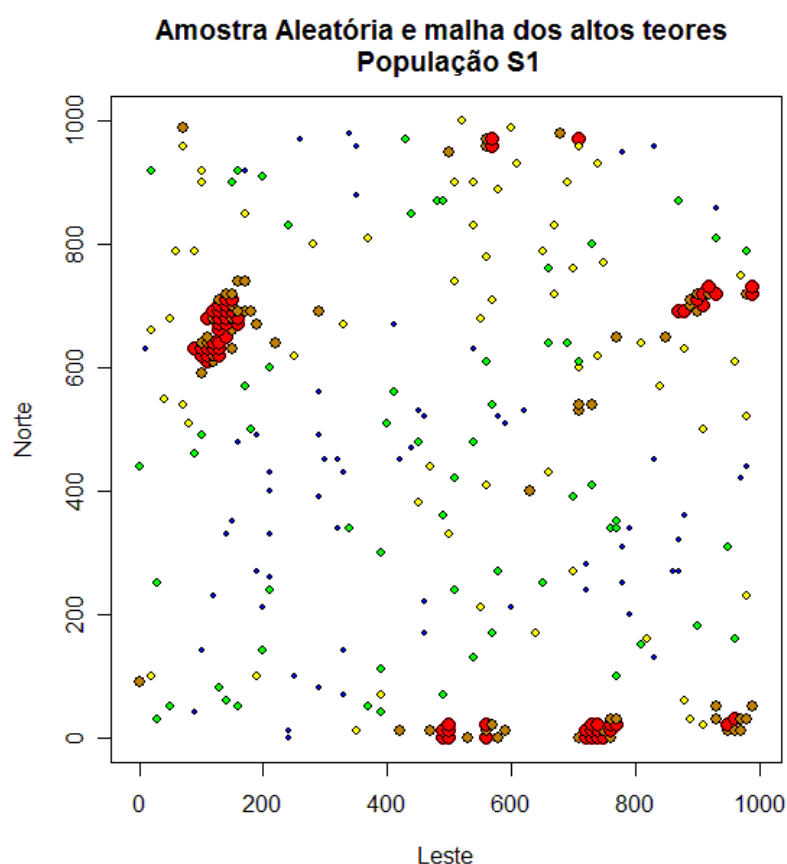


Figura 18. Amostra aleatória. Mapa de distribuição dos pontos de amostragem ao longo da região de estudo. Observam-se os *clusters* dos valores mais altos onde foram estabelecidas malhas regulares de amostragem. O tamanho e cor dos pontos representam classes de valores da variável-alvo. Os valores nos eixos Norte e Leste são expressos em metros.

4.3 PROCEDIMENTO PARA A INTERPOLAÇÃO DOS DADOS

Para a interpolação dos dados amostrais por krigagem ordinária usamos a função `krige` do pacote `gstat` (Pebesma & Wesseling, 1998) do ambiente de programação **R**. Devemos fornecer para esta função: a base de dados (valores e coordenadas dos pontos amostrais), o modelo de covariograma ajustado, o número máximo de pontos amostrais na vizinhança e a distância máxima a partir de cada ponto a ser interpolado além da qual os pontos amostrais não são considerados. Na seção 2.3 vimos que o conjunto dos pontos onde se calcula a estimativa é escolhido em uma malha regular. Assim, devemos fornecer também à função `krige` as coordenadas dos pontos de predição.

A resposta da função `krige` é uma tabela com as coordenadas dos pontos de predição, os valores estimados e as respectivas variâncias de interpolação. Para o cálculo do volume, converteremos a coluna dos valores estimados para a forma matricial.

4.3.1 Vizinhança local

Uma etapa importante no trabalho de interpolação é a definição da vizinhança local, ou seja, os pontos de amostragem que serão efetivamente utilizados pelos métodos de interpolação. Os critérios para seleção de pontos amostrais vizinhos a um ponto a estimar, bem como o número de pontos amostrais a serem utilizados devem ser estabelecidos no início do processo. Essa escolha deve ser feita de tal modo que garanta uma boa amostragem espacial e evite agrupamentos de pontos. Escolhido o critério para a seleção de amostras de furos vizinhos, deve-se definir o número de amostras a ser utilizado. Este não deve ser excessivamente pequeno, com o risco de a interpolação resultar em valor semelhante ou muito correlacionado ao do ponto mais próximo e nem excessivamente grande, com o risco de a interpolação resultar num valor bastante suavizado, perdendo a característica de interpolação local.

Assim, para se evitar agrupamentos de pontos foram estabelecidos critérios de seleção de amostras por quadrantes ou octantes (Yamamoto, 2001), por meio dos

quais os pontos de dados poderiam ser escolhidos, de tal modo que proporcionassem melhor amostragem espacial, isto é, amostras igualmente distribuída em torno do ponto a ser interpolado. Os critérios de seleção de amostras por quadrantes ou octantes dividem a região do ponto a ser interpolado em quatro ou oito setores, respectivamente e selecionam as amostras mais próximas por setor até completar um número desejado de amostras para fins de interpolação (Figura 19). O agrupamento de pontos é um defeito que só ocorre quando os pontos de dados possuem uma distribuição aleatória ou semi-aleatória.

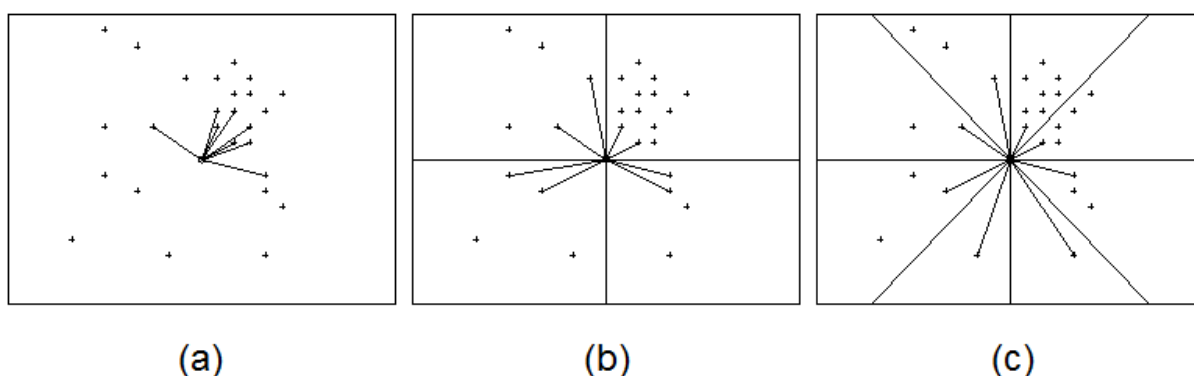


Figura 19. Escolha da vizinhança local a partir de pontos amostrais aleatoriamente distribuídos. Em **(a)** o critério foi o de escolher os oito pontos de dados mais próximos do ponto a ser interpolado. Este critério resulta em um agrupamento de pontos no quadrante nordeste em detrimento dos outros. Em **(b)** usou-se o critério dos quadrantes, escolhendo os dois pontos mais próximos em cada quadrante. Ainda existe uma deficiência de pontos a sul. Em **(c)** usou-se o critério dos octantes, escolhendo o ponto mais próximo em cada octante, o que resulta em uma vizinhança de amostragem representativa em termos do gradiente dos dados. Figura modificada de Yamamoto (2001).

No nosso caso não empregamos os critérios por quadrantes ou octantes, pois a função `kriige` só permite definir, para o entorno de cada ponto a ser interpolado, os números máximo (`nmax`) e mínimo (`nmin`) de pontos amostrais e a distância máxima de busca (`maxdist`). Assim usamos `maxdist=200m`, que é aproximadamente o valor do alcance do semivariograma experimental, e `nmax=inf`, isto é, um número muito grande de modo que para a interpolação são utilizados todos os pontos de dados que se encontram dentro de um raio igual a `maxdist` do entorno do ponto a ser interpolado.

4.3.2 Parâmetros a ajustar

De posse de uma amostra não enviesada e representativa da população visada, o passo seguinte é obter o *variograma teórico* que por sua vez é obtido pelo ajuste dos pontos do semivariograma experimental a uma curva teórica dada por um modelo paramétrico pré-estabelecido. O modelo teórico escolhido, como já foi dito, é o *Matern*. Os parâmetros deste modelo a serem ajustados são: alcance, patamar, efeito pepita e suavização.

Pela análise visual do semivariograma experimental (mostrado na Figura 20) podemos determinar que a variância aleatória da nossa amostra é zero (não apresenta efeito pepita). Com isso, os parâmetros de ajuste são: alcance, patamar e suavização.

5 RESULTADOS

5.1 SEMIVARIOGRAMA EXPERIMENTAL

Para o cálculo do semivariograma experimental consideramos que a análise exploratória das anisotropias já fora implementada e os valores do ângulo azimutal da direção de maior continuidade e da razão de anisotropia da população foram obtidos. Para relembrar, estes valores são, respectivamente, N45°E e 2. Com isso, calcula-se o semivariograma experimental na direção de maior continuidade espacial da variável-alvo. Os valores da variância γ em função da distância h são apresentados na Tabela 1. O código-fonte do programa que calcula este semivariograma é apresentado no Apêndice B.


	np	h	$\gamma(h)$
1	55	14,14213562	0,035097981
2	96	32,73654503	0,070228273
3	116	51,48574904	0,12055388
4	96	71,20232316	0,137179731
5	104	92,3683797	0,354517947
6	55	111,3884205	0,590942096
7	82	129,0248426	0,860427435
8	93	151,2514583	1,054979612
9	90	169,7635278	1,205810091
10	103	189,7894111	0,929773553
11	101	211,746769	0,850220569
12	104	229,8525135	1,109708987
13	100	249,1063954	1,208177898
14	111	269,9216321	1,38601465
15	125	290,7442306	1,394189835
16	121	310,5690068	1,334902199
17	151	330,9471728	1,73540292
18	142	350,0018722	2,147875951
19	149	369,0937722	1,977024595
20	166	390,9255303	1,911325624
21	155	411,2841377	1,90487217
22	163	431,2444734	1,798721087
23	172	449,6887836	1,473842928
24	163	470,0727102	1,146434171
25	213	489,8408035	1,058613642

Tabela 1. Valores do semivariograma experimental obtido a partir da amostra segundo a direção N45°E e com tolerância angular de $\pm 15^\circ$. Definimos 25 “classes de distâncias” ou passos de h e distância máxima de 500 m entre dois pontos como limite para o cálculo da variância. A coluna designada por np traz a quantidade de pares de pontos que entraram no cálculo da variância para cada classe de distância. A largura do intervalo de distância (tolerância de distâncias), igual a duas vezes Δh , foi definida como sendo de 20 m.

A Figura 20 apresenta o semivariograma experimental da amostra. Pode-se perceber o alto grau de continuidade na origem e a inexistência de efeito pepita. O ajuste ao modelo teórico deve ser feito tendo por base apenas os pontos do campo estruturado, pois a distribuição dos pontos no campo aleatório não deve seguir nenhuma lei determinística. A primeira questão que se coloca para quem se propõe a ajustar uma curva teórica sobre os pontos do semivariograma experimental é, portanto, onde termina o campo estruturado e começa o campo aleatório. Ou em outras palavras, qual é o alcance do semivariograma. E a que patamar o variograma se estabiliza. As respostas a estas perguntas determinam a qualidade do ajuste e, neste momento, pesa a experiência do intérprete e o seu conhecimento da geologia da região e a sua familiaridade com a base de dados.

Após a definição do alcance e do patamar, tradicionalmente feita por tentativas e erro e baseando-se em critérios altamente subjetivos, o ajuste da curva teórica é costumeiramente feito com o emprego do método dos mínimos quadrados ou o método dos mínimos quadrados ponderados ou mesmo o ajuste “a sentimento”. O uso de computação para a escolha de valores dos parâmetros também é freqüente.

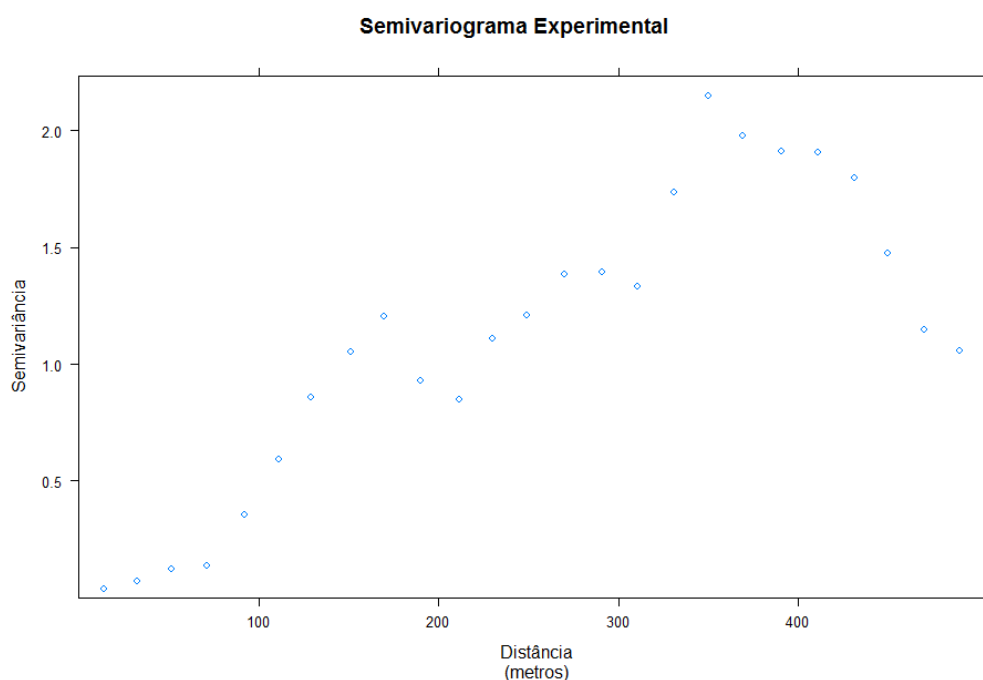


Figura 20. Semivariograma experimental com 25 pontos calculados. Visualmente podemos notar a grande continuidade na origem e a ausência de efeito pepita. Entretanto, a escolha dos valores ótimos de alcance e patamar não é tarefa tão simples quando se emprega apenas a análise visual dos pontos.

5.2 AJUSTE POR ESQUEMA DE PLANEJAMENTO DE EXPERIMENTOS

Na seção anterior explicitamos o caráter altamente subjetivo da abordagem tradicional à tarefa de se buscar os valores dos parâmetros que proporcionam o melhor ajuste teórico aos pontos do semivariograma experimental. Além disso, se esse procedimento iterativo de tentativa e erro não é feito de forma sistemática e se não segue a nenhum planejamento criterioso, ele pode ser muito pouco eficiente e recair em redundâncias ou em ações irrelevantes para o cumprimento da meta. Mesmo após a obtenção de um ajuste final satisfatório, a tarefa pode ter sido desnecessariamente exaustiva.

A nossa proposta é a de formular um procedimento criterioso e sistemático para a escolha dos valores ótimos dos parâmetros de ajuste do variograma teórico baseado no conceito de Planejamento de Experimentos (*Experiment Design*).

O primeiro passo é definir para cada parâmetro de ajuste valores mínimo, intermediário e máximo que sejam razoáveis. No caso do patamar e do alcance, podemos fazer isto com o auxílio do semivariograma experimental por simples análise visual dos pontos. Escolhemos para o alcance os valores 150, 250 e 350. Para o patamar escolhemos os valores 0.7, 1.2 e 1.7. Para o terceiro parâmetro, κ , a escolha dos valores requer uma análise mais cuidadosa. Para isso, construímos um gráfico que mostra como o volume da superfície interpolada varia em função do valor de κ . Este gráfico é apresentado na Figura 21. Escolhemos para o parâmetro de suavização os valores 0.5, 1 e 3 que referem-se à faixa onde os valores de volume possuem ordem de grandeza compatível com os dados.

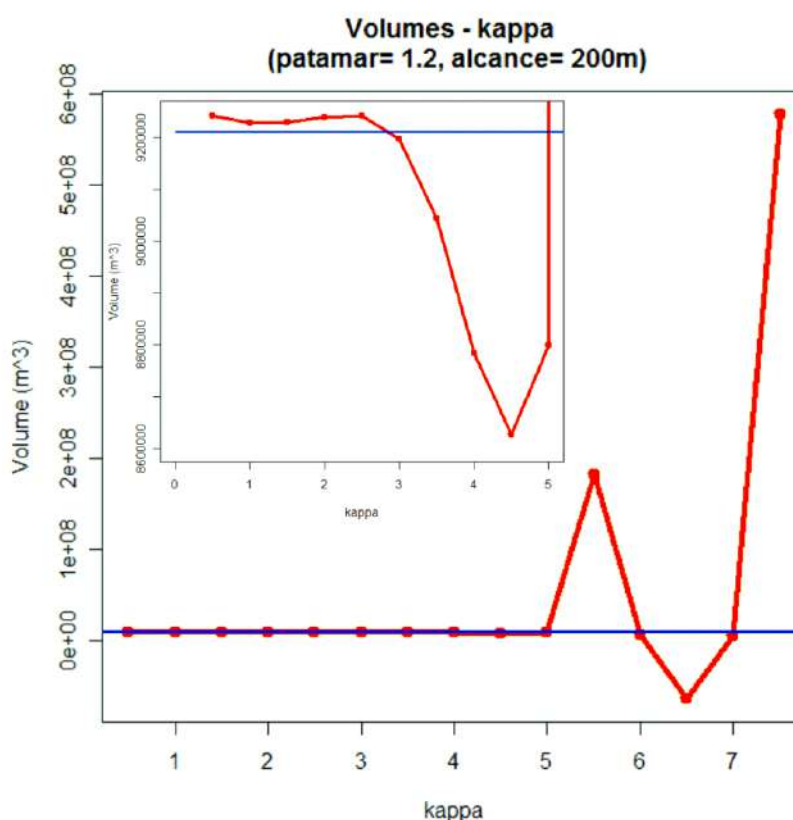


Figura 21. Comportamento do volume da superfície interpolada em função do valor do parâmetro de suavização (κ). A curva em cor vermelha representa o volume da superfície interpolada enquanto que a curva azul indica o volume verdadeiro. No detalhe está uma ampliação do trecho do gráfico entre os valores de κ zero e cinco. É possível ver neste gráfico que a ordem de grandeza do valor do volume e de sua taxa de variação fica fora do domínio dos dados para κ maior do que três.

A Tabela 2 apresenta os valores mínimo, intermediário e máximo escolhidos para o alcance, patamar e κ . Estes valores são usados nos experimentos como base para a definição dos valores ótimos destes parâmetros.


	mínimo	intermediário	máximo
Alcance (A)	150 m	250 m	350 m
Patamar (P)	0.7 m ²	1.2 m ²	1.7 m ²
Suavização (κ)	0.5	1.0	3.0

Tabela 2. Valores mínimo, intermediário e máximo considerados razoáveis escolhidos para o alcance, patamar e suavização do modelo teórico de variograma.

Tendo definido os valores limites de cada parâmetro, somos agora capazes de estudar a influência de cada um desses parâmetros no resultado final, ou seja, queremos saber qual conjunto de valores fornece o melhor ajuste do semivariograma, tendo em vista a otimização da variável de controle. Para isso, empregamos um procedimento sistemático, eficiente e bem estruturado em bases de um esquema de Planejamento de Experimentos (*Experiment Design*).

O procedimento consiste em efetuar uma corrida com as combinações entre os parâmetros, ou um arranjo fatorial, e testar o comportamento do sistema para cada situação. Assim, para os parâmetros A (alcance), P (patamar) e K (suavização), os experimentos realizados são: A (variação de A), P (variação de P), K (variação de K), AP (variação conjunta de A e P), AK (variação conjunta de A e K), PK (variação conjunta de P e K) e APK (variação de todos os parâmetros). Tal técnica permite avaliar não somente os efeitos principais dos parâmetros, como também os efeitos de interação entre os mesmos.

A Figura 22 mostra, para cada experimento, como se comporta a variável de controle (volume) em função dos parâmetros. Conclui-se que o parâmetro K é determinante na obtenção do resultado final. Nesta figura, o símbolo **A-** representa todas as combinações em que A assume seu valor mínimo, **A0** representa todas as combinações em que A assume seu valor intermediário e **A+** representa todas as combinações em que A assume seu maior valor. O símbolo **AP-** representa todas as combinações em que A e P assumem seus valores mínimos simultaneamente, e assim por diante.

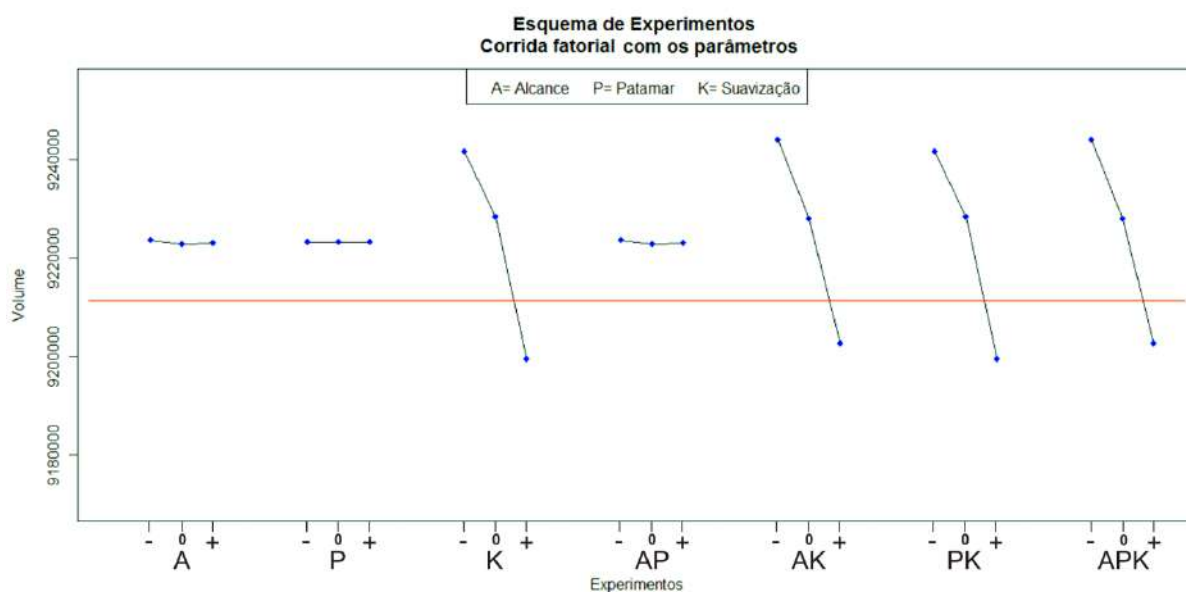


Figura 22. Planejamento fatorial mostrando a contribuição de cada parâmetro no valor final da variável de controle. Os pontos azuis representam o valor do volume da superfície predita a partir de determinada combinação de parâmetros, enquanto que a linha vermelha indica o volume verdadeiro.

Os valores ótimos dos parâmetros podem ser obtidos do estudo do arranjo fatorial de experimentos ilustrado na Figura 22 com a comparação entre o volume predito e o volume verdadeiro. A variação de P dentro dos limites estabelecidos não tem influência alguma sobre o resultado, portanto podemos arbitrar um valor para P dentro dos limites preestabelecidos sem prejuízo da meta. O Parâmetro A exerce sutil influência sobre o volume calculado de modo que podemos fixa tanto para A como para P os seus valores intermediários. O volume calculado se aproxima do volume verdadeiro quando K está próximo de 2. Desse modo, os valores ótimos são $A = 250$ m, $P = 1,2$ m² e $K = 2$.

Com os valores dos parâmetros definidos, plotamos a superfície interpolada por krigagem ordinária que possui mínima diferença de volume em relação ao volume verdadeiro.

A Figura 23 apresenta o histograma e curva de frequência acumulada para a superfície interpolada, que aparece na forma de mapa com código de cores na Figura 24.

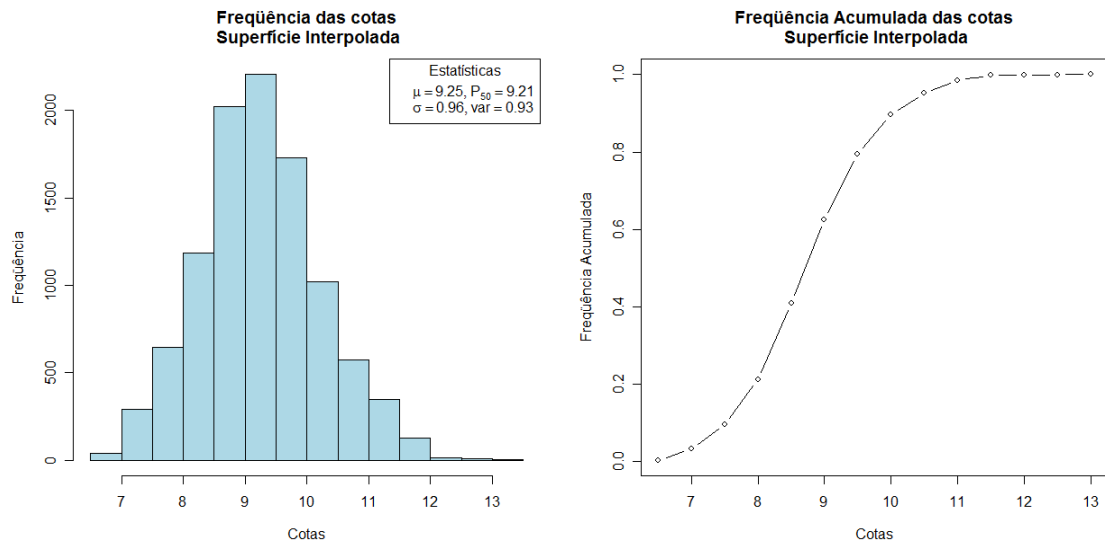


Figura 23. Histograma e curva de frequência acumulada referentes à superfície predita.

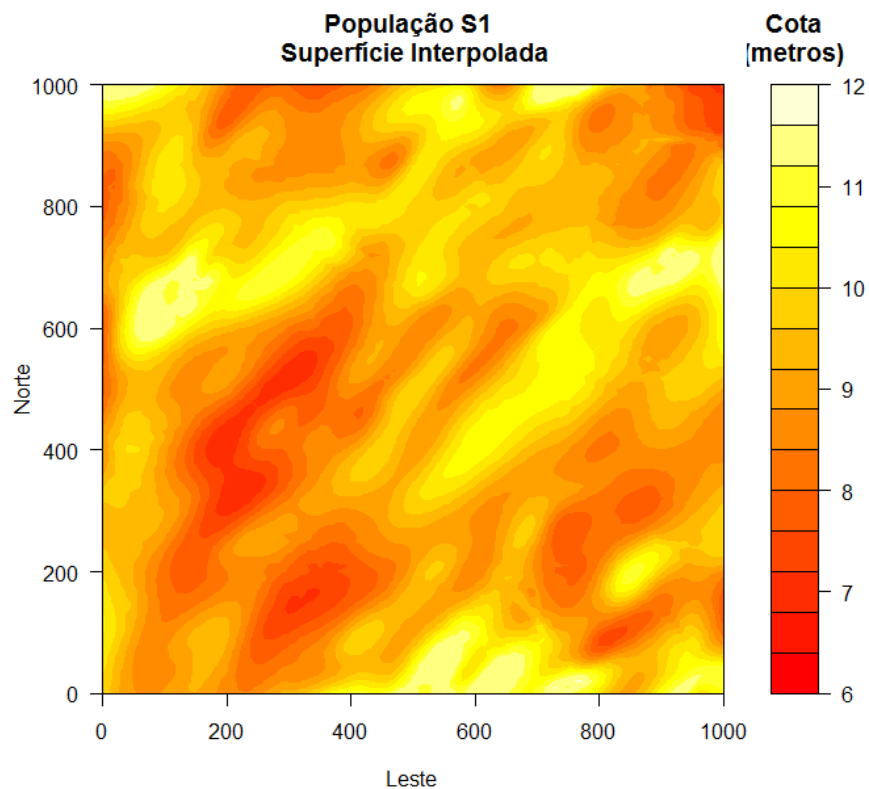


Figura 24. Superfície predita usando-se os parâmetros estabelecidos pelo esquema de planejamento de experimentos. Código de cores refere-se às cotas em relação a um nível de referência arbitrário. A gridagem usada possui resolução espacial (*pixel*) de 10 m para coincidir com a resolução da população visada.

A Figura 25 faz uma comparação entre a população visada (esquerda) e a “população predita” (direita). Percebe-se que os valores da população predita são mais suaves do que os valores originais. Isto se deve ao fato de que a resolução amostral (distância média entre os pontos de dados) é maior do que a variabilidade espacial da população visada (tamanho da menor estrutura da população na escala de trabalho). Então, a variabilidade mínima que a amostra é capaz de “enxergar” é da ordem de grandeza da resolução amostral. Neste trabalho, a resolução espacial da população é de 10 metros enquanto que a resolução espacial média da amostra é de 60 metros.

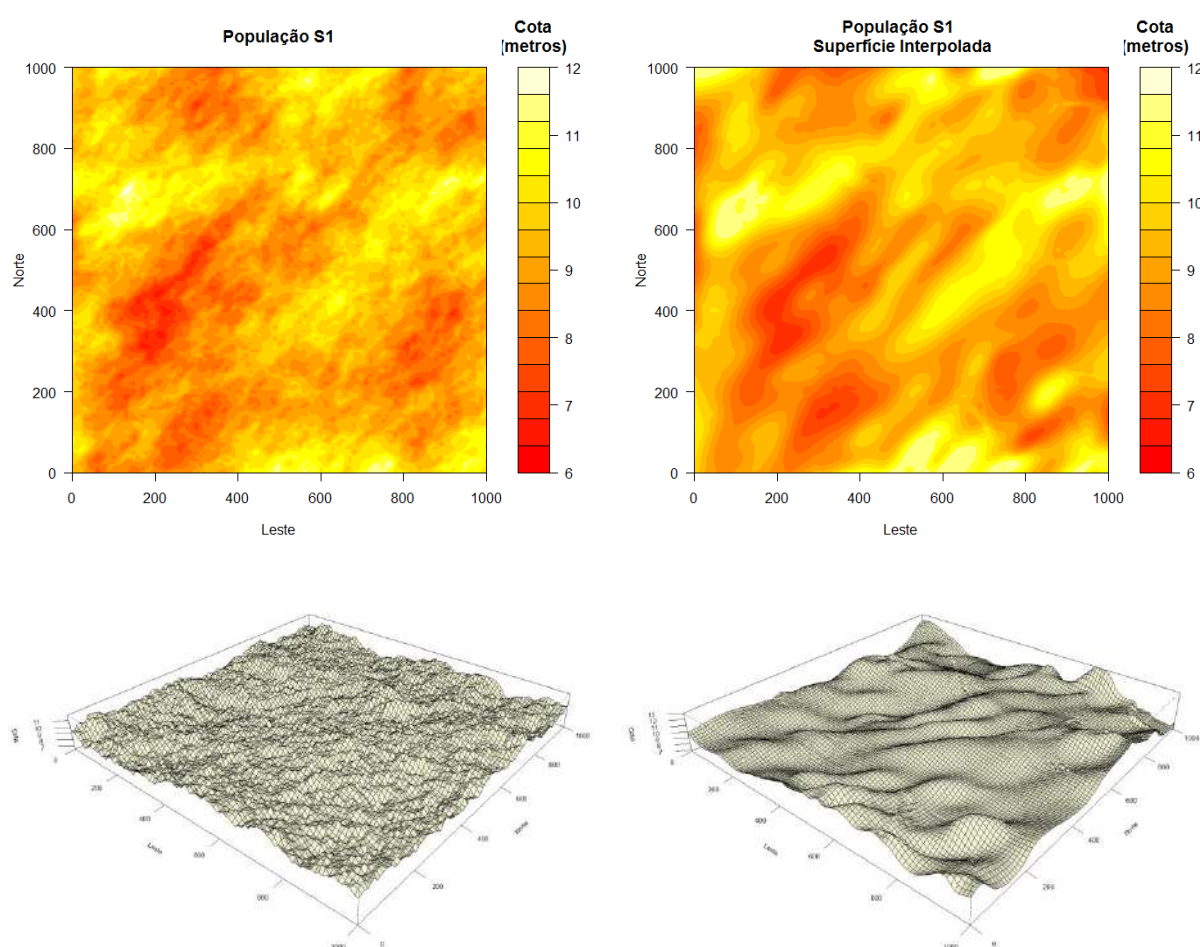


Figura 25. Comparação entre a população visada (esquerda) e a “população predita” (direita).

Então, a qualidade da predição é sempre condicionada à qualidade da amostra, mas podemos maximizar o ganho de qualidade com a amostra que temos. Em relação aos procedimentos de ajuste “tradicionais” dos parâmetros, que são centrados no ajuste sobre o semivariograma experimental, o procedimento baseado

em esquemas de planejamento de experimentos, que são centrados no interpolador de krigagem, nos permite avaliar o ajuste diretamente pela variável de controle escolhida que no caso do atual trabalho é o volume (ou as cotas). A Figura 26 traz dois mapas de comparação entre as cotas preditas e as cotas verdadeiras, um para o caso de ajuste “tradicional” e o outro para o caso de ajuste por esquema de planejamento de experimentos visando o volume.

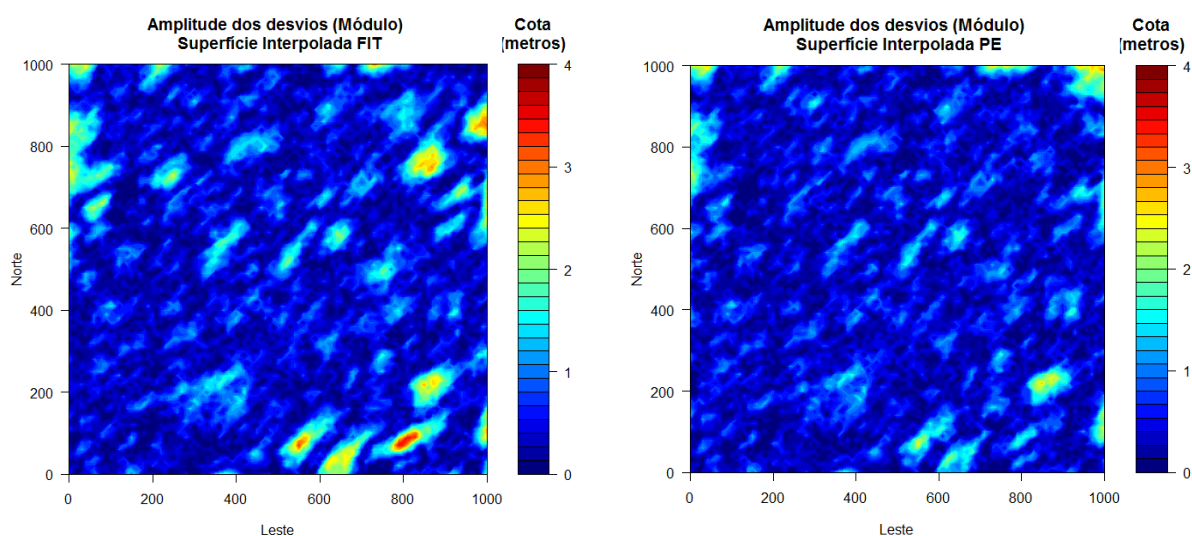


Figura 26. Nestes mapas, o código de cores representa o módulo da diferença de cota entre a superfície original e a superfície interpolada. À esquerda, a escolha dos parâmetros do variograma foi feita do modo “tradicional”, isto é, ajuste “a sentimento” sobre o variograma com auxílio da função `fit.variogram`. À direita, a escolha foi feita pelo esquema de planejamento de experimento. Este último ajuste, de fato, minimiza as diferenças de cota.

De fato, ambos os resultados são razoáveis, o que era de se esperar. Entretanto, como já dissemos, os procedimentos “tradicionais” tendem a ser exaustivos e a exigir muita habilidade do intérprete. Já o procedimento que utiliza esquemas de planejamento de experimentos é conciso, sistemático, objetivo e fornece o melhor ajuste visando um determinado parâmetro de controle, que neste trabalho foi o volume, mas que pode ser qualquer outro parâmetro, por exemplo, a Variância de Interpolação.

6 CONCLUSÃO

O ajuste dos parâmetros do modelo teórico *Matern* ao semivariograma experimental (alcance, patamar e suavização) pode ser feito através de procedimento que utiliza esquema de planejamento de experimentos e visando uma discrepância mínima das cotas em relação aos valores verdadeiros. Além disso, este procedimento é conciso e objetivo e pode ser empregado no lugar dos procedimentos que chamamos de “tradicionais” (aqueles que se baseiam no ajuste de uma curva sobre os pontos do semivariograma experimental). O parâmetro de controle pode ser qualquer quantidade escolhida em função do interesse da pesquisa.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, C.F.P., 2001. A Variação do Suporte Amostral e as Mudanças na Estrutura do Semivariograma. Dissertação de Mestrado, Instituto de Matemática Universidade Federal do Rio de Janeiro, Rio de Janeiro, 104p.
- Almeida, C.F.P., Braga, L.P.V., Bettini, C., 2001. A Variação do Suporte Amostral e as Mudanças na Estrutura do Semivariograma. Anais do III Workshop Brasileiro de Geoinformática - GEOINFO 2001. v. Único. p. 9-16.
- Annels, A.E., 1991. Mineral Deposit Evaluation: a Practical Approach, Chapman&Hall.
- Arantes, D., Mackenzie, B.W., 1995. A Posição Competitiva do Brasil na Mineração de Ouro; Estudos de Política e Economia Mineral, nº.7, DNPM.
- Borges, C.C., 2005. A Topologia: Considerações Teóricas e Implicações para o Ensino da Matemática. Caderno de Física da UFES, 03 (2): 15-35.
- Bussab, W.O., Morettin, P.A.. 2004. Estatística Básica, 5ª ed., São Paulo : Ed. Saraiva, ISBN: 85-02-03497-9, 523p.
- Chaves, H.A.F., Resende, J., Porto Alegre, H.K., 1988. Geoestatística Aplicada à Pesquisa do Xisto, Formação Irati, São Mateus do Sul, Paraná; B. Geoci. PETROBRAS, Rio de Janeiro, 2(2/4): 209-215.
- Cressie, N.A.C., 1993. Statistics for Spatial Data, Revised Edition. John Wiley & Sons, ISBN: 0-471-00255-0, 900p.
- David, M., 1982. Geostatistical Ore Reserve Estimation. Developments in Geomathematics 2. Elsevier Co. ISBN: 0-444-41532-7. 360p.
- Davis, J.C., 1986. Statistics and Data Analysis in Geology, 2nd edition. John Wiley & Sons, 650p.

- Guilbert, J.M., Park Jr, C.F., 1997. The Geology of Ore Deposits, New York: W.H. Freeman and Company. ISBN: 0-7167-1456-6, 985p.
- Guimarães, A.O., 2002. Modelagem Geológica e Geoestatística de Reservatórios da Bacia de Campos, RJ. Dissertação de Mestrado, Departamento de Geologia Universidade Federal do Rio de Janeiro, Rio de Janeiro, 101p.
- Harvard, R., 2005. Gaussian Markov Random Fields: Theory and Applications; Boca Raton: Chapman & Hall/CRC, 263p.
- John, J.A., 1977. Experiments: Design and Analysis, London: C. Griffin, 296p.
- Landim, P.M.B., 2003. Análise Estatística de Dados Geológicos, 2ª ed., São Paulo: Ed. UNESP, ISBN: 85-7139-504-7, 253p.
- Litch, O.A.B., 1998. Prospecção Geoquímica: Princípios, técnicas e métodos, Rio de Janeiro: CPRM, 236p.
- Monteiro, M.C., 2005. Procedimentos Geoestatísticos para Quantificar a Incerteza em Modelos Estáticos de Reservatórios Petrolíferos. Tese de doutoramento em geologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Murrel, P., 2006. R Graphics; Boca Raton: Chapman & Hall/CRC, 301p.
- Pebesma, E.J., Wesseling, C.G., 1998. Gstat: a Program for Geostatistical Modelling, Prediction and Simulation. Computers & Geosciences Vol. 24, No. 1: 17-31.
- R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0.
- Ribeiro JR., P.J., Diggle, P.J., 2001. geoR: A Package for Geostatistical Analysis. R-NEWS Vol 1, No 2 :15-18. ISSN 1609-3631.
- Schmidt, A.M., Nobre, A.A., Ferreira, G.S., 2002. Alguns Aspectos da Modelagem de Dados Espacialmente Referenciados. Revista Brasileira de Estatística, **63**, 220, 59-88.
- Schobbenhaus, C., Silva Coelho, C.E., 1988. Principais Depósitos Minerais do Brasil. Metais Básicos não Ferrosos, Ouro e Alumínio, vol. III, DNPM.
- Yamamoto, J.K., 2001. Avaliação e Classificação de Reservas Minerais, São Paulo : Ed. da USP, ISBN: 85-314-0626-9, 226p.

Apêndice A. O Projeto R

R é um sistema de "programação de dados" idealizado para trabalhos em análise e modelagem estatística e que fornece de forma integrada, entre outras coisas, uma linguagem de programação orientada a objetos, recursos para geração de gráficos de alto nível, interface para outras linguagens de programação, etc. **R** é, portanto, uma suíte integrada de recursos computacionais para manipulação de dados, cálculos e exibição de gráficos. **R** é um programa de código aberto desenvolvido a partir do programa comercial S, que foi criado no final da década de 1980 nos laboratórios da AT&T. O Projeto R começou em 1995 com os trabalhos de Robert Gentleman e Ross Ihaka, do Departamento de Estatística da Universidade de Auckland e rapidamente se difundiu nos meios de pesquisa em estatística. Atualmente é mantido pela **R Core-development Team**, que é um time internacional de desenvolvedores voluntários que continuamente desenvolvem documentação de ajuda, tutoriais e pacotes funcionais que acrescentam bastantes potencialidades à já poderosa versão base do **R**. A página na Internet do Projeto **R**, <http://www.r-project.org> possui todas as informações sobre o **R**, além de disponibilizar gratuitamente o programa **R** e seus pacotes.

Alguns dos benefícios de se usar o **R**:

- Programa de código aberto (gratuito). Fácil de obter via *download* e de instalar. Funciona perfeitamente em várias plataformas, tais como Windows, UNIX e Macintosh;
- Facilidade e extrema liberdade de manipular e estocar dados. Versatilidade na entrada e saída de dados, podendo obter dados de arquivos de texto ASCII, planilha do MS-Excel ou da Internet;
- Grande variedade de operadores para cálculos sobre objetos, principalmente vetores e matrizes;
- Coleção vasta, coerente e integrada de funções internas e pacotes para análise de dados, várias voltadas para estatística e geoestatística;
- Opções variadas para apresentação gráfica de resultados e dados em tela ou armazenado em arquivo de imagem no computador;

- Linguagem de programação simples, mas muito versátil cuja sintaxe é superficialmente similar à da linguagem C, mas com a diferença de que as funções são capazes de receber vetores, matrizes ou outras funções como argumento. Isto é freqüentemente útil, por exemplo, em modelagem estatística;
- Excelente conjunto interno de arquivos de ajuda.

Algumas carências do **R**:

- Interface gráfica limitada. O programa S, ao contrário, possui ótima interface gráfica;
- Não há suporte comercial ao produto. Existem, no entanto, *mailing list* e fóruns internacionais (acessíveis a partir da página do projeto **R**) que minimizam a falta de suporte técnico;
- Praticamente todas as ações dentro do ambiente **R** são executadas via linha de comando, o que exige do usuário inicial um rápido aprendizado da sintaxe da linguagem, mesmo para executar as mais simples tarefas.

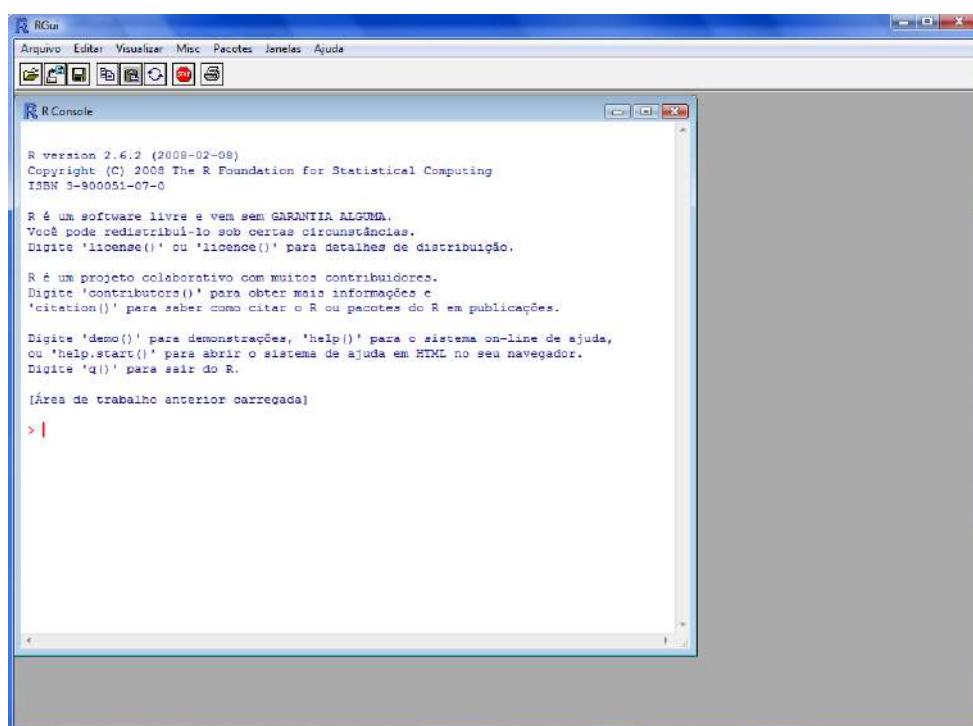


Figura 27. Tela inicial do **R**.

Fundamentos da Linguagem **R**

O **R** é uma linguagem baseada em objetos. Isto quer dizer que tudo o que nós vamos usar no **R** está guardado na memória do computador sob a forma de um

objeto. Todos os objetos em **R** têm um nome associado e podem armazenar diferentes tipos de coisas (números, texto, vetores, matrizes, expressões, etc.).

Para armazenar algo num objeto usamos o operador de atribuição. Este operador consiste num sinal `<` seguido por um sinal `-`, como se vê no exemplo apresentado em seguida, em que guardamos o número 0.55 no objeto que resolvemos chamar de *coef.frac*:

```
> coef.frac <- .55
```

A linguagem **R** também é *case-sensitive*, isto é, faz distinção entre letras maiúsculas e minúsculas. Com isso, um objeto de nome *Taxa* é diferente de outro de nome *taxa*.

O **R** possui 24 tipos de variáveis, como por exemplo, “*char*”, “*logical*”, “*integer*”, “*double*”, “*character*”, etc. As variáveis são armazenadas nos *objetos*. O objeto mais básico do **R** para guardar dados é o vetor. Um vetor é uma estrutura de dados que permite armazenar um conjunto de valores do mesmo tipo (por exemplo números) sob um mesmo nome. No **R**, um número é considerado um vetor de comprimento igual a um.

As funções do **R** definem um conjunto de operações sobre um objeto. O **R** possui um vasto conjunto de funções internas e de funções fornecidas por pacotes, mas o programador também pode criar as suas próprias funções. Assim, para um dado vetor *x*, as instruções:

```
> range(x)
> x*x
> mean(x)
> var(x)
```

Retornam, respectivamente, o valor máximo e mínimo de *x*, um vetor em que cada elemento é igual ao quadrado do respectivo elemento de *x*, a média dos valores de *x* e a variância de *x*. A possibilidade de se usar vetores como argumento de funções torna os programas menos pesados do que os escritos em outras linguagens, pois descarta a necessidade de cálculos iterativos (esta característica é, por vezes, referida como *aritmética vetorial*).

Apêndice B. Códigos-fonte dos programas criados neste trabalho

Aqui apresento todos os códigos-fonte que foram criados para os objetivos deste trabalho, a fim de mostrar como se operou com a linguagem **R** na busca dos resultados e para serem conferidos por aqueles cujo interesse pela sintaxe e pela flexibilidade de trabalho possibilitada pela linguagem por ventura tenha sido despertado em maior grau. De posse destes códigos, é possível reproduzir todas as etapas realizadas neste trabalho a partir de qualquer computador com o **R** previamente instalado, bastando para isto transferir todas as linhas para o console do **R**. Os pacotes usados foram: *geoR*, *RandomFields* e *gstat*.

CÓDIGO 1) Geração da população

```
#####
#####
###      POPULAÇÃO 1: NORMAL - MEAN= 10; STDEV= .4      ###
### NUGGET= 0; PSILL= 1; RANGE= 0.25; KAPPA= 0.5; ANISO= (45°, 2) ###
###      SÉRGIO M. FERREIRA      ###
###      DEGEO/UFRJ - 20/06/2008      ###
#####
###      Geração da população usando grf{geoR}      ###
#####

#####
#Tamanho da população e malha

nx<- 100          #Pontos na direcao x, no intervalo [0,nx]
ny<- nx          #Pontos na direcao y, no intervalo [0,ny]
n<- nx*ny        #Número de nós para a simulação
e<- 10           #Fator de escala para metros
xi<- 0
xf<- e*nx        #Metros
yi<- xi
yf<- e*ny

#####
#Parâmetros input para a simulação

nugget<- 0        #Variância do erro (pepita)
aniso<- c(pi/4, 2) #Anisotropia: direção (em radianos) e intensidade
kappa<- 0.5       #Parâmetro de suavização adicional usado por "matern"
cov<- c(1,0.25)   #Parâmetros de covariância: patamar e alcance
mean<- 10

#####
#Geração da população sintética

f<- grf(n, grid="reg", nx=nx+1, ny=ny+1, nsim=1, cov.model="matern",
      cov.pars=cov, kappa=kappa, nugget=nugget, lambda=1,
      aniso.pars=aniso, mean=mean, method="RF", RF=TRUE, messages=FALSE)

f$coords<- f$coords*xf
```

```
#####
#Visualização dos dados em tela

zlim<- range(f$data, finite=TRUE)
nlevels<- 16 #numero de niveis desejado + 1
levels<- seq(floor(zlim[1]), ceiling(zlim[2]), length=nlevels)
x <- seq(xi, xf, e)
y <- seq(yi, yf, e)
z<- matrix(f$data, nx+1,ny+1)

windows()
image(x, y, z, col=heat.colors(100), xlab="Leste", ylab="Norte", axes=TRUE)

windows()
filled.contour(x, y, z, zlim=zlim, levels=levels, filled=TRUE,
               col=heat.colors(nlevels-1),
               plot.title = title(main = "População S1",
                                   xlab="Leste", ylab="Norte"),
               key.title=title(main="Cota\n(metros)"))

windows()
persp(x, y, z, xlab = "Leste", ylab = "Norte", zlab = "Cota",
      theta = 40, phi = 30, scale = FALSE, expand = 20,
      col = "lightyellow", border = NA, ltheta=45, lphi=20, shade=0.6,
      box = TRUE, axes = TRUE, nticks = 5, ticktype = "detailed")

#####
### POPULAÇÃO REPRESENTADA EM HISTOGRAMA E CURVA DE FREQ. ACUM. ###
#####

## Guardando em data frame os parâmetros da população S1 ##

distrib.param<- data.frame(nugget=nugget, psill=cov[1], range=cov[2],
                            aniso.angle=aniso[1], aniso.ratio=aniso[2],
                            cov.model="matern", kappa=kappa,
                            mean=mean(f$data), P50=quantile(f$data,
                                                                prob=0.5, type=9),
                            std.dev=sd(f$data), var=var(f$data))

#####
# Plotando a distribuição de frequências da população

h<- hist(f$data, breaks = "Sturges", plot = FALSE)

windows()
{plot(h, xlim = range(h$breaks), col = "lightblue",
     main = paste("Frequência das cotas\n População S1"), xlab = "Cotas",
     ylab = "Frequência")

legend("topright", title= "Distribuição normal",
      expression(list(mu == 9.22, P[50] == 9.24),
                  list(sigma^2 == 0.85, var == 0.73)))
}

## Plotando a freq. acum. dos percentis pelos intervalos do histograma ##

fal<- 0
fa<- vector()
for (i in 1:length(h$counts)){
  fal<- fal + h$counts[i]
  fa<- append(fa, values=fal, after = length(fa))
}
```

```

windows()
{plot(x=h$breaks[1:(length(h$breaks)-1)], y=fa/length(f$data),
      main = paste("Frequência Acumulada das cotas\n População S1"),
      xlab = "Cotas", ylab = "Frequência Acumulada", type="b")

legend("right", title= "Distribuição normal",
      expression(list(mu == 10.0, P[50] == 9.2),
      list(sigma^2 == 4.1, var == 16.2)))
}

#####
### EXECUTE AGORA AS LINHAS DO ARQUIVO "SALVA_S1.TXT" PARA ###
### SALVAR TODOS OS OBJETOS (GRÁFICOS E TABELAS) EM DISCO. ###
### * * * * * ###
### FIM DO PROGRAMA ###
#####

```

CÓDIGO 2) Campanha de amostragem aleatória

```

#####

#####
### POPULAÇÃO 1: NORMAL - MEAN= 10; STDEV= .4 ###
### NUGGET= 0; PSILL= 1; RANGE= 0.25; KAPPA= 0.5; ANISO= (45°, 2) ###
### SÉRGIO M. FERREIRA ###
### DEGEO/UFRJ - 20/06/2008 ###
#####
### Geração de amostra aleatória ###
#####

#####
#SUBSETS

#Criar o subset contendo os 100 valores mais altos:
biggest<- 100
prob<- 1-(biggest/length(f$data))
Q1<- quantile(f$data, prob=prob, type=9)
subset1<- f[f$data>=Q1,]

#Criar o subset com valores acima do 3o quartil, menos os "biggest":
Q2<- quantile(f$data, prob=.75, type=9)
subset2<- f[f$data>=Q2 & f$data<Q1,]

#Criar o subset com valores abaixo do 3o quartil:
subset3<- f[f$data<Q2,]

#####
#Amostragem dos subsets e geração da Amostra

#####
### Subset1 - Amostra 1 (malha fina regular nos valores mais altos)

x.am<- subset1$x
y.am<- subset1$y
data.am<- subset1$data

#am1#
am1<- subset1

```

```
#####
### Subset2 - Amostra 2 (valores intermediários)

frac2<- .02      #Fração a ser amostrada em [0,1]
size2<- round(frac2*length(subset2$data))
amt2<- sample(subset2$data, size=size2, replace=FALSE)

search.2a<- match(subset2$data, amt2, nomatch=0)
search.2b<- vector()
for(i in 1:length(search.2a)){
  if(search.2a[i]>0){
    search.2b<- append(search.2b, values= i, after=length(search.2b))
  }
}

for(i in search.2b){
  x.am<- append(x.am, values= subset2$x[i], after=length(x.am))
  y.am<- append(y.am, values= subset2$y[i], after=length(y.am))
  data.am<- append(data.am, values= subset2$data[i], after=length(data.am))
}

#am2#
x2<- y2<- data2<- vector()
for(i in search.2b){
  x2<- append(x2, values= subset2$x[i], after=length(x2))
  y2<- append(y2, values= subset2$y[i], after=length(y2))
  data2<- append(data2, values= subset2$data[i], after=length(data2))
}

am2<- data.frame(x=x2, y=y2, data=data2)

#####
### Subset3 - Amostra 3 (Valores mais baixos)

frac3<- .02      #Fração a ser amostrada em [0,1]
size3<- round(frac3*length(subset3$data))
amt3<- sample(subset3$data, size=size3, replace=FALSE)

search.3a<- match(subset3$data, amt3, nomatch=0)
search.3b<- vector()
for(i in 1:length(search.3a)){
  if(search.3a[i]>0){
    search.3b<- append(search.3b, values= i, after=length(search.3b))
  }
}

for(i in search.3b){
  x.am<- append(x.am, values= subset3$x[i], after=length(x.am))
  y.am<- append(y.am, values= subset3$y[i], after=length(y.am))
  data.am<- append(data.am, values= subset3$data[i], after=length(data.am))
}

#am3#
x3<- y3<- data3<- vector()
for(i in search.3b){
  x3<- append(x3, values= subset3$x[i], after=length(x3))
  y3<- append(y3, values= subset3$y[i], after=length(y3))
  data3<- append(data3, values= subset3$data[i], after=length(data3))
}

am3<- data.frame(x=x3, y=y3, data=data3)

#####
#AMOSTRA
```


CÓDIGO 3) Cálculo do Semivariograma Experimental

```
#####

#####
###          POPULAÇÃO 1: NORMAL - MEAN= 10; STDEV= .4          ###
### NUGGET= 0; PSILL= 1; RANGE= 0.25; KAPPA= 0.5; ANISO= (45°, 2)  ###
###                      SÉRGIO M. FERREIRA                      ###
###                      DEGEO/UFRJ - 20/06/2008                  ###
#####
###          Cálculo do Semivariograma Experimental          ###
#####

#####
#VARIograma EXPERIMENTAL

windows()
cutoff<- (xf-xi)/2          #Raio máximo de abrangência
np<- 25                     #Número de pontos no semivariograma experimental

vexp<- variogram(data~l, location=~x+y, data=amostra, alpha=45, tol.hor=15,
                 cutoff=cutoff, width=cutoff/np)

plot(vexp, main = "Semivariograma Experimental",
      xlab="Distância\n(metros)", ylab="Semivariância")

#####
#SALVANDO A TABELA DE DADOS E GRÁFICO

bmp(file="C:/Arquivos/Ajuste/S1/SVexp/vexp.bmp", width=15, height=10,
     units="cm", res= 160)
plot(vexp, main = "Semivariograma Experimental",
      xlab="Distância\n(metros)", ylab="Semivariância")
dev.off()

write.table(vexp,file="C:/Arquivos/ Ajuste/S1/SVexp/vexp.txt",
            append=FALSE, quote=FALSE, sep=" ",
            row.names=TRUE, col.names=TRUE)

#####
###          * * * * *          ###
###                      FIM DO PROGRAMA                      ###
###          * * * * *          ###
#####
```

CÓDIGO 4) Análise variacional do parâmetro de suavização (κ)

```
#####

#####
###          POPULAÇÃO 1: NORMAL - MEAN= 10; STDEV= .4          ###
### NUGGET= 0; PSILL= 1; RANGE= 0.25; KAPPA= 0.5; ANISO= (45°, 2)  ###
###                      SÉRGIO M. FERREIRA                      ###
###                      DEGEO/UFRJ - 20/06/2008                  ###
#####
###          Análise Variacional do parâ. de suavização (kappa)  ###
#####

#####
```

```
#VAMOS BUSCAR UMA FAIXA DE VALORES RAZOAVEIS PARA KAPPA, MANTENDO OS
#OUTROS PARAMETROS CONSTANTES EM: PSILL= 1.2; ALCANCE= 200m E NUGGET= 0
```

```
#####
#Atribuição de valores às constantes do sistema
var1<- 1.2                                #patamar
var2<- 200                                #alcance
var3<- seq(from=0.5, to=7.5, by=0.5)      #kappa length=100
nugget<- 0                                #Efeito Pepita
aniso<- c(45, 0.5)                        #Anisotropia
```

```
#####
#Definição da malha de predição
pred<- expand.grid(x=x, y=y)
```

```
#####
#Definição de funções úteis
```

```
#####
#                               FUNÇÃO 1: calcula os volumes                               #
#####
vol<- function(z, e=10){
  nrow<- nrow(z)
  ncol<- ncol(z)
  Sb<- e^2
```

```

#Calculo da matriz m2
v1<- c(1,rep(2,nrow-2),1)
v2<- 2*v1
v<- c(v1,rep(v2,ncol-2),v1)
m2<- matrix(v,nrow,ncol)
```

```

#Calculo do volume
m<- z*m2
a<- sum(m)
volume <- (Sb * a)/4
volume
}
```

```
#####
#                               FUNÇÃO 2:                               #
#      Gera um vetor com os volumes para cada valor de kappa           #
#      *      *      *      *      *      *      *      *      *      #
#      O argumento é um vetor com os valores de kappa                   #
#####
vols<- function(arg){
  N<- length(arg)
  volumes<- vector()
  for(i in 1:N){
    vmodel<- vgm(psill=var1, model="Mat", range=var2,
                 nugget=nugget, kappa=arg[i], anis=aniso)

    kg<- krige(formula=data~1, locations=~x+y, data=amostra,
               newdata=pred, model=vmodel, maxdist=(xf/5))

    z.pred<- matrix(kg$var1.pred, nx+1,ny+1)

    vl<- vol(z.pred)
    volumes<- append(volumes, values=vl, after=length(volumes))
  }
  volumes
}
```

```
#####
#Volume verdadeiro (da população S1)
vv<- vol(z)

#####
#Estudo da variação de kappa
vkappa<- vols(var3)

#####
#Armazenar os dados em um data frame
kappa<- data.frame(kappa=var3, volume=vkappa)

#####
#SALVANDO A TABELA DE DADOS E GRÁFICO

bmp(file = "C:/Arquivos/S1/Ajuste_kappa/kappa1.bmp",
    width=10, height=10, units="cm", res= 160)

plot(kappa, type = "o", col = "red", lwd=4,
     main = "Volumes - kappa \n (patamar= 1.2, alcance= 200m)",
     xlab="kappa", ylab="Volume (m^3)")
lines(x=c(0,8), y=c(vv,vv), col="blue", lwd=2)
dev.off()

bmp(file = "C:/Arquivos/S1/Ajuste_kappa/kappa2.bmp",
    width=10, height=10, units="cm", res= 160)

plot(kappa, type = "o", col = "red", lwd=4, xlim=c(0,5),
     ylim=c(8600000,9245000),
     main = "Volumes - kappa \n (patamar= 1.2, alcance= 200m)",
     xlab="kappa", ylab="Volume (m^3)")
lines(x=c(0,5.5), y=c(vv,vv), col="blue", lwd=2)
dev.off()

write.table(kappa, file="C:/Arquivos/S1/Ajuste_kappa/kappa.txt",
            append=FALSE, quote=FALSE, sep=" ", row.names=TRUE,
            col.names=TRUE)

#####
###                                FIM DO PROGRAMA                                ###
###                                * * * * * * * * * *                                ###
#####
```

CÓDIGO 5) Esquema de planejamento de experimento e predição da superfície por krigagem ordinária

```
#####
#####
###          POPULAÇÃO 1: NORMAL - MEAN= 10; STDEV= .4          ###
### NUGGET= 0; PSILL= 1; RANGE= 0.25; KAPPA= 0.5; ANISO= (45°, 2) ###
###                      SÉRGIO M. FERREIRA                      ###
###                      DEGEO/UFRJ - 20/06/2008                  ###
#####
### Ajuste do variograma com esquema de Planejam. de Experimentos ###
#####
```



```

#K.0
K.0<- expand.grid(A=A, P=P, K.0=K[2])          #Alcance, Patamar, Kappa (médio)
vols.K.0<- vols(K.0)                          #Vetor com todos os volumes usados nos cálculos
v.K.0<- mean(vols.K.0)                        #Volume médio para o médio Kappa

#K.M
K.M<- expand.grid(A=A, P=P, K.M=K[3])          #Alcance, Patamar, Kappa (alto)
vols.K.M<- vols(K.M)                          #Vetor com todos os volumes usados nos cálculos
v.K.M<- mean(vols.K.M)                        #Volume médio para o alto Kappa

#####
#Alcance(A) e Patamar(P)

#AP.m
AP.m<- expand.grid(A.m=A[1], P.m=P[1], K=K)      #A (baixo), P (baixo)
vols.AP.m<- vols(AP.m)                        #Vetor com todos os volumes usados nos cálculos
v.AP.m<- mean(vols.AP.m)                      #Volume médio para os mínimos alcance e patamar

#AP.0
AP.0<- expand.grid(A.0=A[2], P.0=P[2], K=K)      #A (médio), P (médio)
vols.AP.0<- vols(AP.0)                        #Vetor com todos os volumes usados nos cálculos
v.AP.0<- mean(vols.AP.0)                      #Volume médio para os médios alcance e patamar

#AP.M
AP.M<- expand.grid(A.M=A[3], P.M=P[3], K=K)      #A(alto), P(alto)
vols.AP.M<- vols(AP.M)                        #Vetor com todos os volumes usados nos cálculos
v.AP.M<- mean(vols.AP.M)                      #Volume médio para os altos alcance e patamar
#####
#Alcance(A) e Kappa(K)

#AK.m
AK.m<- expand.grid(A.m=A[1], P=P, K.m=K[1])      #A(baixo), Kappa (baixo)
vols.AK.m<- vols(AK.m)                        #Vetor com todos os volumes usados nos cálculos
v.AK.m<- mean(vols.AK.m)                      #Volume médio para os mínimos alcance e Kappa

#AK.0
AK.0<- expand.grid(A.0=A[2], P=P, K.0=K[2])      #A(médio), Kappa (médio)
vols.AK.0<- vols(AK.0)                        #Vetor com todos os volumes usados nos cálculos
v.AK.0<- mean(vols.AK.0)                      #Volume médio para os médios alcance e Kappa

#AK.M
AK.M<- expand.grid(A.M=A[3], P=P, K.M=K[3])      #A(alto), Kappa (alto)
vols.AK.M<- vols(AK.M)                        #Vetor com todos os volumes usados nos cálculos
v.AK.M<- mean(vols.AK.M)                      #Volume médio para os altos alcance e Kappa

#####
#Patamar(P) e Kappa(K)

#PK.m
PK.m<- expand.grid(A=A, P.m=P[1], K.m=K[1])      #P(baixo), Kappa (baixo)
vols.PK.m<- vols(PK.m)                        #Vetor com todos os volumes usados nos cálculos
v.PK.m<- mean(vols.PK.m)                      #Volume médio para os mínimos patamar e Kappa

#PK.0
PK.0<- expand.grid(A=A, P.0=P[2], K.0=K[2])      #P(médio), Kappa (médio)
vols.PK.0<- vols(PK.0)                        #Vetor com todos os volumes usados nos cálculos
v.PK.0<- mean(vols.PK.0)                      #Volume médio para os médios patamar e Kappa

#PK.M
PK.M<- expand.grid(A=A, P.M=P[3], K.M=K[3])      #P(alto), Kappa (alto)
vols.PK.M<- vols(PK.M)                        #Vetor com todos os volumes usados nos cálculos
v.PK.M<- mean(vols.PK.M)                      #Volume médio para os altos patamar e Kappa

```

```
#####
#Alcance(A), Patamar(P) e Kappa(K)

#APK.m
APK.m<- expand.grid(A.m=A[1], P.m=P[1], K.m=K[1])      #A, P, Kappa (baixos)
v.APK.m<- vols(APK.m)      #Vol médio para os mínimos alcance, patamar e Kappa

#APK.0
APK.0<- expand.grid(A.0=A[2], P.0=P[2], K.0=K[2])      #A, P, Kappa (médios)
v.APK.0<- vols(APK.0)      #Vol médio para os médios alcance, patamar e Kappa

#APK.M
APK.M<- expand.grid(A.M=A[3], P.M=P[3], K.M=K[3])      #A, P, Kappa (altos)
v.APK.M<- vols(APK.M)      #Vol médio para os altos alcance, patamar e Kappa

Sys.time() - t

#####
#Unindo todos os volumes médios em um data frame

vpe<- data.frame(A=c(v.A.m, v.A.0, v.A.M), P=c(v.P.m, v.P.0, v.P.M),
                 K=c(v.K.m, v.K.0, v.K.M), AP=c(v.AP.m, v.AP.0, v.AP.M),
                 AK=c(v.AK.m, v.AK.0, v.AK.M), PK=c(v.PK.m, v.PK.0,
                 v.PK.M), APK=c(v.APK.m, v.APK.0, v.APK.M),
                 Nivel=c("baixo", "medio", "alto"))

#####
#Gerando a superfície interpolada com os parâmetros finais escolhidos

A.final<- A[2]                #Alcance final
P.final<- P[2]                #Patamar final
K.final<- 2                   #Kappa final

vmmodel<- vgm(psill=P.final, model="Mat", range=A.final,
              nugget=nugget, kappa=K.final, anis=aniso)

kg.final<- krige(formula=data~1, locations=~x+y, data=amostra,
                 newdata=pred, model=vmmodel, maxdist=(xf/5))

z.pred.final<- matrix(kg.final$var1.pred, nx+1,ny+1)

#####
#Guardando em data frame os parâmetros da população S1

distrib.param<- data.frame(mean=mean(kg.final$var1.pred),
                            P50=quantile(kg.final$var1.pred, prob=0.5,
                            type=9), std.dev=sd(kg.final$var1.pred),
                            var=var(kg.final$var1.pred))

#####
#Plotando a distribuição de frequências da população

h<- hist(kg.final$var1.pred, breaks = "Sturges", plot = FALSE)

windows()
{plot(h, xlim = range(h$breaks), col = "lightblue",
      main = paste("Frequência das cotas\n Superfície Interpolada"),
      xlab = "Cotas", ylab = "Frequência")}

legend("topright", title= "Estatísticas", expression(list(mu == 9.25, P[50]
== 9.21), list(sigma == 0.96, var == 0.93)))
}
```

```
## Plotando a frequência acumulada ##

fal<- 0
fa<- vector()
for (i in 1:length(h$counts)){
  fal<- fal + h$counts[i]
  fa<- append(fa, values=fal, after = length(fa))
}

windows()
{plot(x=h$breaks[1:(length(h$breaks)-1)], y=fa/length(kg.final$var1.pred),
      main = paste("Frequência Acumulada das cotas\n Superfície Interpolada"), xlab = "Cotas", ylab = "Frequência Acumulada", type="b")}

legend("topright", title= "Estatísticas",
       expression(list(mu == 9.25, P[50] == 9.21),
                    list(sigma == 0.96, var == 0.93)))
}

#####
#Plotando a superfície interpolada em curvas de nível e em perspectiva

zlim<- range(f$data, finite=TRUE)
nlevels<- 16 #numero de niveis desejado + 1
levels<- seq(floor(zlim[1]), ceiling(zlim[2]), length=nlevels)

filled.contour(x, y, z.pred.final, xlim=zlim, levels=levels, filled=TRUE,
               col=heat.colors(nlevels-1),
               plot.title = title(main = "População S1 \n Superfície Interpolada", xlab="Leste", ylab="Norte"),
               key.title=title(main="Cota\n(metros)"))

persp(x, y, z.pred.final, xlab = "Leste", ylab = "Norte", zlab = "Cota",
      theta = 40, phi = 30, scale = FALSE, expand = 20, col = "lightyellow", border = NULL, ltheta = 45, lphi = 20, shade = 0.6, box = TRUE, axes = TRUE, nticks = 5, ticktype = "detailed")

#####
### EXECUTE AGORA AS LINHAS DO ARQUIVO "SALVA_S1_PExp.TXT" ###
### PARA SALVAR TODOS OS OBJETOS (GRÁFICOS E TABELAS) EM DISCO. ###
### * * * * * ###
### FIM DO PROGRAMA ###
#####
```

CÓDIGO 6) Comparação dos módulos das diferenças das cotas

```
#####
#####
###          POPULAÇÃO 1: NORMAL - MEAN= 10; STDEV= .4          ###
### NUGGET= 0; PSILL= 1; RANGE= 0.25; KAPPA= 0.5; ANISO= (45°, 2) ###
###                      SÉRGIO M. FERREIRA                      ###
###                      DEGEO/UFRJ - 20/06/20008                  ###
#####
###                      Módulo das Diferenças das cotas          ###
#####
```

```
#####
#Comparação das cotas. Parâmetros ajustados por P.E.

zx<- z-z.pred.final
zz<- sqrt(zx*zx)
zlim<- range(zz, finite=TRUE)
nlevels<- 31 #numero de niveis desejado + 1
levels<- seq(floor(zlim[1]), ceiling(zlim[2]), length=nlevels)

jet.colors <-
  colorRampPalette(c("#00007F", "blue", "#007FFF", "cyan",
                    "#7FFF7F", "yellow", "#FF7F00", "red", "#7F0000"))

bmp(file = "C:/Arquivos/Campo IV/Simulacoes/6_Conclusao/Fit/zdif_PE.bmp",
     width=11, height=10, units="cm", res= 160)
filled.contour(x, y, zz, zlim=zlim, levels=levels, filled=TRUE,
               col=jet.colors(nlevels-1), plot.title = title(main =
                    "Amplitude dos desvios (Módulo) \n Superfície Interpolada PE",
                    xlab="Leste", ylab="Norte"), key.title=title(main="Cota\n
                    (metros)"))
dev.off()

#####
#Comparação das cotas. Parâmetros ajustados do modo "tradicional"

vmodel<- vgm(psill=1, model="Mat", range=cutoff, nugget=nugget,
             kappa=.5, anis=aniso)

ajuste<- fit.variogram(vexp, vmodel)

zx<- z-z.pred
zz<- sqrt(zx*zx)
zlim<- range(zz, finite=TRUE)
nlevels<- 31 #numero de niveis desejado + 1
levels<- seq(floor(zlim[1]), ceiling(zlim[2]), length=nlevels)

bmp(file = "C:/Arquivos/Campo IV/Simulacoes/6_Conclusao/Fit/zdif_FIT.bmp",
     width=11, height=10, units="cm", res= 160)

filled.contour(x, y, zz, zlim=zlim, levels=levels, filled=TRUE,
               col=jet.colors(nlevels-1), plot.title = title(main =
                    "Amplitude dos desvios (Módulo) \n Superfície Interpolada FIT",
                    xlab="Leste", ylab="Norte"), key.title=title(main="Cota\n
                    (metros)"))
dev.off()

#####
###          * * * * *          ###
###          FIM DO PROGRAMA          ###
###          * * * * *          ###
#####
```